

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau

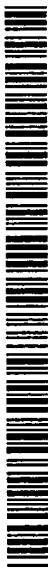


(43) International Publication Date  
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number  
**WO 01/77350 A2**

- (51) International Patent Classification<sup>7</sup>: C12N 15/79      (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (21) International Application Number: PCT/US01/11436
- (22) International Filing Date: 4 April 2001 (04.04.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
09/545,574                    7 April 2000 (07.04.2000)    US
- (71) Applicant (*for all designated States except US*): LARGE SCALE BIOLOGY CORPORATION [US/US]; 3333 Vaca Valley Parkway, Vacaville, CA 95688 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): PALMER, Kenneth, E. [ZA/US]; 707 West Monte Vista Avenue, Vacaville, CA 95688 (US). POGUE, Gregory, P. [US/US]; 419 Trillick Court, Vacaville, CA 95688 (US).
- (74) Agent: HALLUIN, Albert, P., Howrey Simon Arnold & White, 301 Ravenswood Avenue, Box 34, Menlo Park, CA 94025 (US).



Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**WO 01/77350 A2**

(54) Title: COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION

(57) Abstract: The present invention provides an eukaryotic recombinant vector suited for bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the transgene in an eukaryotic cell. The invention vectors are particularly suited for mediating gene silencing in a variety of biological systems. The present invention also provides host cells and transgenic plants comprising the invention vectors. Further provided by the invention are methods of inhibiting expression of an endogenous gene present in an eukaryotic cell. Also included is a method of identifying a biological function(s) of an endogenous gene of interest in an eukaryotic cell by selectively inhibiting the expression of the endogenous gene.

**COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION**

5

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Patent Application  
10 09/545,574, filed April 7, 2000, pending, which is hereby incorporated herein by  
reference in its entirety.

STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER  
FEDERALLY SPONSORED RESEARCH

15

Not applicable.

TECHNICAL FIELD

This invention is in the field of genetic analysis. Specifically, the invention  
relates to the generation of a eukaryotic vector that allows bi-directional  
transcription of a transgene to yield both sense and antisense RNA transcripts from  
20 the same transgene. The compositions and methods embodied in the present  
invention are particularly useful for targeted inhibition of gene expression in a  
eukaryotic cell.

25

BACKGROUND OF THE INVENTION

The structure and biological behavior of a cell is determined by the pattern of  
gene expression within that cell at a given time. Perturbations of gene expression  
have long been acknowledged to account for a vast number of diseases including,  
numerous forms of cancer, vascular diseases, neuronal and endocrine diseases.  
30 Abnormal expression patterns, in form of amplification, deletion, gene  
rearrangements, and loss or gain of function mutations, are now known to lead to  
aberrant behavior of a disease cell. Aberrant gene expression has also been noted as  
a defense mechanism of certain organisms to ward off the threat of pathogens.

One of the major challenges of genetic engineering has been to regulate the expression of targeted genes that are implicated in a wide diversity of physiological responses. While overexpression of an exogenously introduced transgene in a eukaryotic cell is relatively straightforward, targeted inhibition of specific genes has been more difficult to achieve. Traditional approaches for suppressing gene expression, including site-directed gene disruption, antisense RNA or co-suppressor injection, require complex genetic manipulations or heavy dosages of suppressors that often exceeds the toxicity tolerance level of the host cell.

Recently, a new technique, "double-stranded RNA interference" has emerged in the study of gene silencing. Several research groups have demonstrated a marked inhibition of a specific nuclear gene expression in a wide range of eukaryotes by introduction into cells of dsRNA fragments that bear sequence homology with the nuclear gene. For instance, Fire et al. (1998) *Nature* 395: 854 reported the success of gene-specific interference in *C. elegans* that was mediated by ingested *E. coli* carrying a prokaryotic vector capable of producing both sense and antisense RNAs of the selected *C. elegans* genes. Misquitta et al. demonstrated the targeted disruption of *nautilus* gene in *Drosophila melanogaster* by injecting into the Drosophila embryo multiple copies of *nautilus* dsRNA. See Misquitta et al. (1999) *PNAS U.S.A.* 96:1451-1456. Studies by Ngô et al. (1998) *Proc. Natl. Acad. of Sci. U.S.A.*, 96:1451-1456 confirmed that dsRNA interference also occurs in certain protozoan species. Earlier studies by Cogoni et al. and Hamilton et al. suggested that formation of dsRNA play a pivotal role in gene silencing in fungi *Neurospora crassa* and other plants. See Cogoni et al. (1999) *Nature* 399: 166-169; Hamilton et al. (1999) *Science* 286: 950-952; and Waterhouse et al. (1999) *PNAS U.S.A.* 95: 13959-13964. More recent investigations by Wargelius et al. revealed that this phenomenon is also conserved in vertebrates such as the zebrafish. Wargelius et al. *Biochem. Biophys. Res. Commun.* 263: 156-161.

Current techniques for achieving RNA mediated gene silencing include: (a) use of prokaryotic vectors capable of transcribing both sense and antisense RNA (Fire et al. (1998) *Nature* 395: 854; (b) *in vitro* transcription of individual strands of a selected gene followed by annealing the transcribed sense and antisense RNAs (see, e.g. Misquitta et al. (1999) *PNAS U.S.A.* 96:1451-1456); and possibly (c) viruses induced gene silencing (see, e.g. Angell et al. (1997) *EMBO Journal* 16:

3675-3684; Angell et al. (1999) *Plant Journal* 20: 357-362). However, these methods bear a number of intrinsic limitations. First, none of these methods employs gene delivery vehicles that are applicable for consistent and persistent inhibition of gene expression in a eukaryote. Second, these existing methods do not necessarily result in production of a substantially homogenous population of dsRNAs. Notably, the *in vitro* preparation of double-stranded RNAs by transcribing and annealing sense RNA transcripts to antisense transcripts is time consuming, labor intensive, and not amenable for mass production or high-throughput analyses.

5 Thus, there remains a considerable need for compositions and methods to effect dsRNA-mediated gene silencing. An ideal reagent would be a self-replicating vector that is (a) capable of autonomous replication and expression of a selected transgene in a eukaryotic cell; and (b) capable of yielding both sense and antisense RNA transcripts from the same transgene, so as to effect production of dsRNA transcripts in a eukaryotic host cell. The present invention satisfies these needs and provides related advantages as well.

10

15

20

25

30

SUMMARY OF THE INVENTION

A principal aspect of the present invention is the design of a eukaryotic recombinant vector to effect gene silencing in a eukaryotic cell that is susceptible to dsRNA-mediated reduction of gene expression. Such a vector allows bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the same transgene in a eukaryotic cell. While not being bound to any one theory, the production of dsRNAs induces transcriptional and/or post-transcriptional gene silencing in the host cell. Accordingly, the present invention provides a recombinant vector having the following unique characteristics: it comprises a viral replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene fragment in a eukaryotic host cell.

In one aspect of this embodiment, each of the overlapping transcription units of the vector comprises a promoter and a terminator that are arranged in one of the configurations shown in Figure 2(a)-(d). The promoter can be constitutive or

inducible; it can be active in all tissues and cell types of an organism or operative only in selected tissues (i.e. tissue-specific).

In another aspect, the recombinant vector comprises a viral replicon that is derived from a DNA virus. Such DNA viruses can be selected from the group consisting of *Geminivirus*, *Caulimoviridae*, *Badnaviridae*, *Circoviridae*, *Circinoviridae*, *Parvoviridae*, *Papovaviridae*, *Polyomaviridae*, *Adenoviridae*, *Herpesviridae*, *Poxviridae*, *Iridoviridae*, *Baculoviridae*, *Hepadnaviridae*, *Retroviridae*, *Gyrovirus*, *Nanovirus*, and African Swine Fever virus.

In yet another aspect, the subject vector is capable of autonomous replication in a eukaryotic cell.

In still another aspect, the subject vector is capable of inhibiting expression of genes endogenous to a eukaryotic host cell. Non-limiting representative eukaryotic cells whose gene expression can be inhibited upon introduction of the subject vectors are fungi, yeast cells, plant cells, insect, avian, mammalian or other animal cells. Preferably, the vectors effect a reduced expression of an endogenous gene that is substantially homologous to the transgene contained in the overlapping transcription units of the vectors. More preferably, delivery of the vectors into a suitable host cell results in a phenotypic change of the host cell. In certain preferred embodiments, the endogenous gene is native to the host cell. The endogenous gene can also be heterologous to the host cell. In some embodiments, the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa. The transgene carried in the vector can be a nucleotide sequence that encodes a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, or a chaperon protein.

The present invention also provides host cells transformed with the invention vectors. The present invention further provides a transgenic plant comprising a eukaryotic recombinant vector of the present invention.

Also provided by the present invention is a kit for generating a double-stranded RNA transcript in a eukaryotic cell that contains the subject vectors in suitable packaging.

Further embodied in the present invention is a method of inhibiting expression of an endogenous gene present in a eukaryotic cell. The method involves: (a) providing a eukaryotic recombinant vector containing a transgene

that is substantially homologous to the endogenous gene; (b) introducing the eukaryotic recombinant vector into the eukaryotic cell; and (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene that is contained in the transcription units of the vector, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

Also included in the present invention is a method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene. The method comprises: (a) providing a eukaryotic recombinant vector containing a transgene that is substantially homologous to the endogenous gene; (b) introducing the eukaryotic recombinant vector of (a) into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the eukaryotic recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and (d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell. In essence, the subject methods allow the creation of a transient or more long-term gene-specific knock-out system for analyzing the biological function of any endogenous gene of interest.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of the process for production of dsRNA transcripts by a subject vector containing two overlapping transcription units.

Figure 2 (a)-(d) depict four different configurations of the overlapping transcription units of the subject vectors.

Figure 3 is a schematic representation of an exemplary construct MSVLSB-6.

Figure 4 depicts the nucleotide sequence of the vector pMSVLSB-1 (SEQ ID NO:9) described in Examples 1-2.

Figure 5 depicts the nucleotide sequence of the vector pMSVLSB-2 (SEQ ID NO:10) described in Examples 1-2.

Figure 6 depicts the nucleotide sequence of the vector pMSVLSB-3 (SEQ ID NO:11) described in Examples 1-2.

5 Figure 7 depicts the nucleotide sequence of the vector pMSVLSB-4 (SEQ ID NO:12) described in Examples 1-2.

Figure 8 depicts the nucleotide sequence of the vector pMSVLSB-5 (SEQ ID NO:13) described in Examples 1-2.

10 Figure 9 depicts the nucleotide sequence of the vector pMSVLSB-6 (SEQ ID NO: 14) described in Examples 1-2.

#### MODES FOR CARRYING OUT THE INVENTION

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

#### **General Techniques:**

20 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See, e.g., Matthews, PLANT VIROLOGY, 3<sup>rd</sup> edition (1991); Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2<sup>nd</sup> edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

25 30 As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

**Definitions:**

A "plant cell" refers to the structural and physiological unit of plants, consisting of a protoplast and the cell wall.

5 A "protoplast" is an isolated cell without cell walls, having the potency for regeneration into cell culture, tissue or whole plant.

10 A "host cell" includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in genomic of total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected *in vivo* with a polynucleotide(s) of this invention.

15 The terms "polynucleotide", "nucleotides" and "oligonucleotides" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, 20 introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

25 A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated.

"Genes of a specific developmental origin" refer to genes expressed at certain but not all developmental stages. For instance, a gene may be of embryonic or adult origin depending on the stage during which the gene is expressed.

5 A "disease-associated" or "disease-causing" gene refers to any gene which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-  
10 associated gene also refers to gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at normal or abnormal level.

15 A gene "database" denotes a set of stored data which represent a collection of sequences including nucleotide and peptide sequences, which in turn represent a collection of biological reference materials.

20 As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are  
25 collected referred to as gene product. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in an eukaryotic cell.

25 "Differentially expressed", as applied to nucleotide sequence or polypeptide sequence in a subject, refers to over-expression or under-expression of that sequence when compared to that detected in a control. Underexpression also encompasses absence of expression of a particular sequence as evidenced by the absence of detectable expression in a test subject when compared to a control.

30 "Differential expression" refers to alterations in the abundance or the expression pattern of a gene product.

A "primer" is a short polynucleotide, generally with a free 3' -OH group, that binds to a target or "template" potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target.

The term "hybridize" as applied to a polynucleotide refers to the ability of the polynucleotide to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues in a hybridization reaction. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. The hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

Hybridization can be performed under conditions of different "stringency". Relevant conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and the washing procedure. Higher stringency conditions are those conditions, such as higher temperature and lower sodium ion concentration, which require higher minimum complementarity between hybridizing elements for a stable hybridization complex to form. In general, a low stringency hybridization reaction is carried out at about 40 °C in 10 x SSC or a solution of equivalent ionic strength/temperature. A moderate stringency hybridization is typically performed at about 50 °C in 6 x SSC, and a high stringency hybridization reaction is generally performed at about 60 °C in 1 x SSC.

When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, the reaction is called "annealing" and those polynucleotides are described as "complementary". A double-stranded polynucleotide can be "complementary" or "homologous" to another polynucleotide, if hybridization can occur between one of the strands of the first polynucleotide and the second. "Complementarity" or "homology" (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonding with each other, according to generally accepted base-pairing rules.

In the context of polynucleotides, a "linear sequence" or a "sequence" is an order of nucleotides in a polynucleotide in a 5' to 3' direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the

polynucleotide. A "partial sequence" is a linear sequence of part of a polynucleotide which is known to comprise additional residues in one or both directions.

The terms "cytosolic", "nuclear" and "secreted" as applied to cellular proteins specify the extracellular and/or subcellular location in which the cellular protein is mostly localized. Certain proteins are "chaperons", capable of translocating back and forth between the cytosol and the nucleus of a cell.

A "subject" as used herein refers to a biological entity containing expressed genetic materials. The biological entity is preferably can be plant, animal, or microorganisms including bacteria, viruses, fungi, and protozoa. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

A "control" is an alternative subject or sample used in an experiment for comparison purpose. A control can be "positive" or "negative". For example, where the purpose of the experiment is to detect a differentially expressed transcript or polypeptide in cell or tissue affected by a disease of concern, it is generally preferable to use a positive control (a subject or a sample from a subject, exhibiting such differential expression and syndromes characteristic of that disease), and a negative control (a subject or a sample from a subject lacking the differential expression and clinical syndrome of that disease).

"Heterologous" means derived from a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a promoter removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous promoter.

A "cell line" or "cell culture" denotes bacterial, plant, insect or higher eukaryotic cells grown or maintained *in vitro*. The descendants of a cell may not be completely identical (either morphologically, genotypically, or phenotypically) to the parent cell.

A "vector" is a nucleic acid molecule, preferably self-replicating, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of a DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression vectors that function for transcription and/or translation of the DNA

or RNA. Also included are vectors that provide more than one of the above functions.

5 An "expression vector" is a polynucleotide which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide(s). An "expression system" usually connotes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

10 A "replicon" refers to a polynucleotide comprising an origin of replication (generally referred to as an ori sequence) which allows for replication of the polynucleotide in an appropriate host cell. Examples of replicons include episomes (such as plasmids), as well as chromosomes (such as the nuclear or mitochondrial chromosomes).

15 A "transcription unit" is a DNA segment capable of directing transcription of a gene or fragment thereof. Typically, a transcription unit comprises a promoter operably linked to a gene or a DNA fragment that is to be transcribed, and optionally regulatory sequences located either upstream or downstream of the initiation site or the termination site of the transcribed gene or fragment.

#### Vectors of the present invention

20 A central aspect of the present invention is the design of a recombinant vector suited for bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the transgene in a eukaryotic cell. The invention vectors are particularly suited for mediating nuclear gene silencing in a variety of biological systems. Distinguished from the previously described DNA vectors, the subject vectors have the following unique characteristics: (a) the vector replicates and directs expression of a transgene in a eukaryotic cell; and (b) the vector 25 comprises a replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene in a eukaryotic host cell.

30 Several factors apply to the design of vectors having the above-mentioned characteristics. First, the vector comprises a replicon having an origin of replication (generally referred to as an ori sequence) which permits replication of the vector in a eukaryotic host cell. A preferred replicon is one comprising viral sequences capable

of directing autonomous replication of the vector in an appropriate host cell. Non-limiting examples of viral replicons include sequences derived from DNA viruses such as *Geminivirus*, *Caulimoviridae*, *Badnaviridae*; *Circoviridae*, *Circinoviridae*, *Parvoviridae*, *Papovaviridae*, *Polyomaviridae*, *Adenoviridae*, *Herpesviridae*,  
5 *Poxviridae*, *Iridoviridae*, *Baculoviridae*, *Hepadnaviridae*, *Retroviridae*, *Gyrovirus*, *Nanovirus*, and African Swine Fever virus, or the like. In addition to the replication origin, a replicon typically carries a transcription unit that directs transcription of a transgene or a fragment thereof to yield a plurality of RNA transcripts.

A second consideration in designing the subject vector is to select two  
10 overlapping transcription units. By "overlapping" is meant that the two transcription units directs transcription of both DNA strands of the same transgene to yield a plurality of partially or perfectly double stranded RNA transcripts. The two overlapping transcription units are typically arranged in an opposing orientation so that each unit can drive transcription of one of the complementary strands from the  
15 same transgene, and thus facilitate the generation of double stranded RNA.  
transcripts. Elements within a transcription unit include but are not limited to promoter regions, enhancer regions, repressor binding regions, transcription initiation sites, ribosome binding sites, translation initiation sites, protein encoding regions and introns, and termination sites for transcription and translation. Preferred transcription  
20 units are arranged in a configuration shown in Figure 2(a)-(d).

As used herein, a "promoter" is a DNA region capable under certain conditions of binding RNA polymerase and initiating transcription of a coding region located downstream (in the 3' direction) from the promoter. It can be constitutive or inducible. In general, the promoter sequence is bounded at its 3'  
25 terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

The choice of promoters will largely depend on the host cells in which the vector is introduced. Commonly employed plant promoters include but are not limited those from agrobacterium, nopaline synthase gene, octopine synthase gene,

mannopine synthase, rbcS (small subunit of ribulose bis-phosphate carboxylase). In addition, the promoter sequences may be provided by viral material. Any RNA virus subgenomic promoters described in Dawson et al. Advances in Virus Research, 38:307-342 and WO93/03161 can thus be employed. For animal cells, a variety of robust promoters, both viral and non-viral promoters, are known in the art. Non-limiting representative viral promoters include CMV, the early and late promoters of SV40 virus, promoters of various types of adenoviruses (e.g., adenovirus 2) and adeno-associated viruses. It is also possible, and often desirable, to utilize promoters normally associated with a desired transgene sequence, provided that such control sequences are compatible with the host cell system. See Goeddel et al., Gene Expression Technology Methods in Enzymology Volume 185, Academic Press, San Diego, (1991), Ausubel et al, Protocols in Molecular Biology, Wiley Interscience (1994).

Suitable promoter sequences for other eukaryotic cells such as yeast cells include the promoters for 3-phosphoglycerate kinase, or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

To optimize the yield of double-stranded RNAs formed from the sense and anti-sense strands transcribed by the overlapping units, it is preferable to use two promoters of comparable strength. The relative strength of the promoters can be determined or ascertained by any conventional recombinant techniques and methods exemplified herein. Representative techniques are Northern blot hybridization and DNA array-based technologies. An illustrative promoter pair comprises MSV mp promoter and CaMV 35S RNA promoter.

Where desired, heterologous promoters that are removed from their native coding sequences and operatively linked to a transgene which it is not naturally

found linked, can be used in constructing the invention vectors. As such, any viral promoters described above can be used to drive the transcription of a non-viral transgenes; promoters of one class of genes can be employed to direct transcription of transgenes coding for other related or unrelated classes of proteins. In certain 5 embodiments of the invention, it is preferable to employ inducible promoters to control the transcription of a transgene. A diverse variety of inducible promoters have been described in the art. Promoters of any endogenous genes whose 10 expressions are inducible by internal or external factors can be employed. Factors applicable for transcription induction include but are not limited to hormones, heat shock, oxygen deficiency, light, stress and various chemicals. Commonly employed inducible promoters are  $\beta$ -gal promoter that is activated upon addition of IPTG; hps70 promoter that is inducible by heat shock; and ribulose-1,5-biphosphate carboxylase (RUBISCO) promoter that is regulated by light.

15 Tissue-specific promoters may also be used. A vast diversity of tissue specific promoters have been described and employed by artisans in the field. Representative plant tissue promoters include that of legumin (or other seed storage protein promoters), patatin and the like. Exemplary promoters operative in selective animal tissue include hepatocyte-specific promoters and cardiac muscle specific promoters. Depending on the intended use of the subject vectors, those skilled in the 20 art will know of other suitable tissue-specific promoters applicable for non-constitutive bi-directional transcription.

In constructing the subject vectors, the termination sequences associated with the transgene are also inserted into the 3' end of the sequence desired to be transcribed to provide polyadenylation of the mRNA and/or transcriptional 25 termination signal. The terminator sequence preferably contains one or more transcriptional termination sequences (such as polyadenylation sequences) and may also be lengthened by the inclusion of additional DNA sequence so as to further disrupt transcriptional read-through. Preferred terminator sequences (or termination sites) of the present invention have a gene that is followed by a transcription 30 termination sequence, either its own termination sequence or a heterologous termination sequence. Examples of such termination sequences, including stop codons coupled to various polyadenylation sequences that are known in the art, widely available, and exemplified below. Where the terminator comprises a gene, it

can be advantageous to use a gene which encodes a detectable or selectable marker; thereby providing a means by which the presence and/or absence of the terminator sequence (and therefore the corresponding inactivation and/or activation of the transcription unit) can be detected and/or selected. Alternatively, a terminator may simply be a second promoter, arranged in inverted orientation to the promoter described above.

The terminators and promoters of the two overlapping transcription units may take a variety of configurations. In one aspect, terminators 1 and 2 of the overlapping transcription units are arranged to immediately flank the transgene as shown in Figure 2(a). In another aspect, the two terminators are placed at the 5' end or the 3' end of their respective promoters as depicted in Figure 2(b). In other aspects, terminator 1 and promoter 1 are flanked by terminator 2 and promoter 2 as shown in Figure 2(c), or vice versa (see Figure 2(d)). Any other variations in configuring the two overlapping transcription units that permit bi-directional transcription are encompassed by the present invention.

The transgene transcribed by an invention vector can be any gene expressed in a eukaryotic cell. The selection of transgene is determined largely by the intended purpose of the vector. Where the vector is used to inhibit expression of an endogenous gene present in a host cell, the transgene selected are substantially homologous to the target endogenous gene. In general, substantially homologous nucleotide sequences are at least about 60% identical with each other, after alignment of the homologous regions. Preferably, the sequences are at least about 75% identical; more preferably, they are at least about 80% identical; more preferably, they are at least about 90% identical; still more preferably, the sequences are 95% identical.

Sequence alignment and homology searches are often determined with the aid of computer methods. A variety of software programs are available in the art. Non-limiting examples of these programs are Blast (<http://www.ncbi.nlm.nih.gov/BLAST/>), Fasta (Genetics Computing Group package, Madison, Wisconsin), DNA Star, MegAlign, and GeneJockey. Any sequence databases that contains DNA sequences corresponding to a target gene or a segment thereof can be used for sequence analysis. Commonly employed databases include but are not limited to GenBank, EMBL, DDBJ, PDB, SWISS-PROT, EST,

STS, GSS, and HTGS. Sequence similarity can be discerned by aligning the transgene sequence against a target endogenous gene sequence. Common parameters for determining the extent of homology set forth by one or more of the aforementioned alignment programs include p value and percent sequence identity.

5 P value is the probability that the alignment is produced by chance. For a single alignment, the p value can be calculated according to Karlin et al. (1990) *Proc. Natl. Acad. Sci.* 87: 2264. For multiple alignments, the p value can be calculated using a heuristic approach such as the one programmed in Blast. Percent sequence identity is defined by the ratio of the number of nucleotide matches between the query sequence and the known sequence when the two are optimally aligned. A selected

10 transgene and target endogenous sequences are considered to be substantially homologous when the regions of alignment exhibit the aforementioned range of percentage of identity using Fasta or Blast alignment program with the default settings.

15 Sequence homology can also be determined by functional analyses. A sequence that preserves the functionality of the nucleic acid with which it is being compared is particularly preferred. Functionality may be established by different criteria, such as ability to hybridize with a target polynucleotide, ability to effectively amplify a target sequence to yield a substantially homogenous multiplicity of products, and the ability to extend the 3' end sequence complementary to a target sequence in a nucleotide sequencing reaction.

20 Where desired, the transgene may comprise heterologous sequences that facilitate detection of the expression and purification of the gene product. Examples of such sequences are known in the art and include those encoding reporter proteins, such as  $\beta$ -galactosidase,  $\beta$ -lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Other heterologous sequences that facilitate purification may code for epitopes such as Myc, HA (derived from influenza virus hemagglutinin), His-6, FLAG, glutathione S-transferase (GST), maltose-binding protein (MBP), or the Fc portion of immunoglobulin.

25 30 The target endogenous genes whose expression is to be inhibited encompass native and heterologous genes present in the host cell. "Native" genes are nucleic acid sequences originated from the host cell. Non-limiting illustrative native genes

5 include those encode membrane proteins, cytosolic proteins, secreted proteins, nuclear proteins and chaperon proteins. Heterologous genes are sequences acquired exogenously by the host cell. Exogenous sequences can be either integrated into the host cell genome, or maintained as episomal sequences. An exemplary class of heterologous genes includes pathogenic genes derived from viruses, bacteria, fungi, and protozoa.

10 The endogenous genes suitable for the present invention may also be characterized based on one or more of the following features: ability to induce a phenotypic change in a host cell or organism, species origin, developmental origin, primary structural similarity, involvement in a particular biological process, association with or resistance to a particular disease or disease stage, tissue, sub-tissue or cell-specific expression pattern, and subcellular location of the expressed gene product. In one aspect, the endogenous gene may be any gene expressed in a eukaryote cell, such as a plant cell, animal cell or a yeast cell. In another aspect, the 15 endogenous gene confers a phenotypic characteristic detectable by visual, microscopic, genetic, or chemical means. Within this class of genes, of particular interest are plant genes involved in growth phenotypes, e.g. stunting, hyperbranching, vein banding, ring spot, etching, and those responsible for color characteristics including bleaching and chlorosis. Also, of particular relevance are genes which upon inhibition provide an enhanced resistance to pathogens (e.g. 20 bacteria, fungi, viruses, insects, and protozoa), and resistance to adverse environmental factors (e.g. temperature fluctuation, nutritional deficiency, adverse soil conditions, moisture, dryness, etc.).

25 In another aspect, the endogenous genes are of a specific developmental origin, such as those expressed in an embryo or an adult organism, during ectoderm, mesoderm, or endoderm formation in a multi-cellular animal, or during development of leaves, tubers, bud of a plant. In yet another aspect, the endogenous genes belong to a family of genes, or a sub-family of genes that share primary structural similarities. Structural similarities can be discerned with the aid of computer 30 software described above. Non-limiting examples of gene families include those encoding proteinase, proteinase inhibitors, cell surface receptors, protein kinases (e.g. tyrosine, serine/threonine or histidine kinases), trimeric G-proteins, cytokines, PH-, SH2-, SH3-, PDZ-domain containing proteins, and any of those gene families

published by the Institute for Genomic Research (TIGR), Incyte Pharmaceuticals, Inc., Human Genome Sciences Inc., Monsanto, and PE-Celera.

In yet another aspect, the endogenous genes are involved in a specific biological process, including but not limited to cell cycle regulation, cell differentiation, chemotaxis, apoptosis, cell motility and cytoskeletal rearrangement.

5 In still another aspect, the endogenous genes embodied in the invention are associated with a particular disease or with a specific disease stage. Such genes include but are not limited to those associated with autoimmune diseases, obesity, hypertension, diabetes, neuronal and/or muscular degenerative diseases, cardiac diseases, endocrine disorders, any combinations thereof. In yet still another aspect, 10 the endogenous genes encompass those exhibiting restricted expression patterns. Non-limiting exemplary gene transcripts of this class include those that are not ubiquitously expressed, but rather are differentially expressed in one or more of the plant tissues including leaf, seed, tuber, stems, root, and bud; or expressed in animal body tissues including heart, liver, prostate, lung, kidney, bone marrow, blood, skin, bladder, brain, muscles, nerves, and selected tissues that are affected by various 15 types of cancer (malignant or non-metastatic), affected by cystic fibrosis or polycystic kidney disease. Additional examples of non-ubiquitously expressed genes are those whose gene products are localized to certain subcellular locations:

20 extracellular matrix, nucleus, cytoplasm, cytoskeleton, plasma and/or intracellular membranous structures which include but are not limited to coated pits, Golgi apparatus, endoplasmic reticulum, endosome, lysosome, and mitochondria.

In addition to the above-described elements, the vectors may contain a selectable marker (for example, a gene encoding a protein necessary for the survival 25 or growth of a host cell transformed with the vector), although such a marker gene can be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells into which a selectable gene has been introduced will survive and/or grow under selective conditions. Typical selection genes encode protein(s) 30 that (a) confer resistance to antibiotics or other toxins substances, e.g., ampicillin, neomycin, methotrexate, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media. The choice of the proper marker gene will depend on the host cell, and appropriate genes for different hosts are known in the art.

The vectors embodied in this invention can be obtained using recombinant cloning methods and/or by chemical synthesis. A vast number of recombinant cloning techniques such as PCR, restriction endonuclease digestion and ligation are well known in the art, and need not be described in detail herein. One of skill in the art can also use the sequence data provided herein or that in the public or proprietary databases to obtain a desired vector by any synthetic means available in the art.

**Host cell and transgenic organisms of the present invention:**

10        The invention provides eukaryotic host cells transformed with the recombinant DNA vectors described above. The recombinant vectors containing the transgene of interest can be introduced into a suitable eukaryotic cell by any of a number of appropriate means, including electroporation, transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; and infection (where the vector is coupled to an infectious agent). The choice of introducing vectors will often depend on features of the host cell.

15        For most animal cells, any of the above-mentioned methods is suitable for vector delivery. For plant cells, a variety of techniques derived from these general methods is available in the art. The host cells may be in the form of whole plants, isolated cells or protoplasts. Preferably, the cells are "intact" in that the cell comprises an outer layer of cell wall, typically composed of cellulose for protection and maintaining the rigidity of the plant cell. Illustrative procedures for introducing vectors into plant cells include Agrobacterium-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs and injection into immature embryos. As is evident to one skilled in the art, each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular plant species may not necessarily be the most effective for another plant species.

20        25        30        Agrobacterium *tumefaciens*-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, bypassing the need for regeneration of an intact plant from a protoplast. The use of Agrobacterium-mediated expression vectors to introduce

DNA into plant cells is well known in the art. This technique makes use of a common feature of *Agrobacterium* which colonizes plants by transferring a portion of their DNA (the T-DNA) into a host cell, where it becomes integrated into nuclear DNA. The T-DNA is defined by border sequences which are 25 base pairs long, and any DNA between these border sequences is transferred to the plant cells as well.

5 The insertion of a recombinant plant viral nucleic acid between the T-DNA border sequences results in transfer of the recombinant plant viral nucleic acid to the plant cells, where the recombinant plant viral nucleic acid is replicated, and then spreads systemically through the plant. Agro-infection has been accomplished with potato spindle tuber viroid (PSTV); CaV; and Lazarowitz, S., *Nucl. Acids Res.* 16:229

10 (1988)) digitaria streak virus (Donson *et al.*, *Virology* 162:248 (1988)), wheat dwarf and tomato golden mosaic virus (TGMV). Therefore, agro-infection of a susceptible plant could be accomplished with a virion containing a recombinant plant viral nucleic acid based on the nucleotide sequence of any of the above viruses. Particle bombardment or electroporation or any other methods known in the art may also be used.

15

Because not all plants are natural hosts for *Agrobacterium*, alternative methods such as transformation of protoplasts may be employed to introduce the subject vectors into the host cells. For certain monocots, transformation of the plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments. See, for example, Potrykus *et al.*, *Mol. Gen. Genet.*, 199:167-177 (1985); Fromm *et al.*, *Nature*, 319:791 (1986); Callis *et al.*, *Genes and Development*, 1:1183 (1987). Applicability of these techniques to different plant species may depend upon the feasibility to regenerate that particular plant species from protoplasts.

20

25

In addition to protoplast transformation, particle bombardment is an alternative and convenient technique for delivering the invention vectors into a plant host cell. Specifically, the plant cells may be bombarded with microparticles coated with a plurality of the subject vectors. Bombardment with DNA-coated microparticles has been successfully used to produce stable transformants in both plants and animals (see, for example, Sanford *et al.* (1993) *Methods in Enzymology*, 217:483-509). Microparticles suitable for introducing vectors into a plant cell are

30

typically made of metal, preferably tungsten or gold. These microparticles are available for example, from BioRad (e.g., Bio-Rad's PDS-1000/He). Those skilled in the art will know that the particle bombardment protocol can be optimized for any plant by varying parameters such as He pressure, quantity of coated particles, 5 distance between the macrocarrier and the stopping screen and flying distance from the stopping screen to the target.

Vectors can also be introduced into plants by direct DNA transfer into pollen as described by Zhou et al., *Methods in Enzymology*, 101:433 (1983); Luo et al., *Plant Mol. Biol. Reporter*, 6:165 (1988). Alternatively, the vectors can be injected 10 into reproductive organs of a plant as described by Pena et al., *Nature*, 325:274 (1987).

Other techniques for introducing nucleic acids into a plant cell include:

- (a) Hand Inoculations. Hand inoculations are performed using a neutral pH, low molarity phosphate buffer, with the addition of celite or carborundum 15 (usually about 1%). One to four drops of the preparation is put onto the upper surface of a leaf and gently rubbed.
- (b) Mechanized Inoculations of Plant Beds. Plant bed inoculations are performed by spraying (gas-propelled) the vector solution into a tractor-driven mower while cutting the leaves. Alternatively, the plant bed is 20 mowed and the vector solution sprayed immediately onto the cut leaves.
- (c) High Pressure Spray of Single Leaves. Single plant inoculations can also be performed by spraying the leaves with a narrow, directed spray (50 psi, 6-12 inches from the leaf) containing approximately 1% carborundum in the buffered vector solution.
- 25 (d) Vacuum Infiltration. Inoculations may be accomplished by subjecting a host organism to a substantially vacuum pressure environment in order to facilitate infection.

Once introduced into a suitable host cell, expression of the transgene can be 30 determined using any assay known in the art. For example, the presence of transcribed sense or anti-sense strands of the transgene can be detected and/or quantified by conventional hybridization assays (e.g. Northern blot analysis), amplification procedures (e.g. RT-PCR), SAGE (U.S. Patent No. 5,695,937), and

array-based technologies (see e.g. U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934). In conducting these analytical procedures, it is preferable to induce transcription of one strand of the transgene at a time. As is apparent to one skilled in the art, the simultaneous transcription of both sense and anti-sense strands facilitates formation of double stranded RNA molecules, which may obscure the accurate determination of the levels of sense and anti-sense RNA transcripts.

5 Expression of the transgene can also be determined by examining the protein product. A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked 10 immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and PAGE-SDS.

15 In general, determining the protein level involves (a) providing a biological sample containing polypeptides; and (b) measuring the amount of any immunospecific binding that occurs between an antibody reactive to the transgene product and a component in the sample, in which the amount of immunospecific binding indicates the level of expressed proteins. Antibodies that specifically recognize and bind to the protein products of the transgene are required for 20 immunoassays. These may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra*. and Sambrook et al. (1989) *supra*. The sample of test proteins can be prepared by homogenizing the eukaryotic transformants (e.g. plant cells) or their progenies made therefrom, and optionally solubilizing the test protein using detergents, preferably 25 non-reducing detergents such as triton and digitonin. The binding reaction in which the test proteins are allowed to interact with the detecting antibodies may be performed in solution, or on a solid tissue sample, for example, using tissue sections or solid support that has been immobilized with the test proteins. The formation of the complex can be detected by a number of techniques known in the art. For 30 example, the antibodies may be supplied with a label and unreacted antibodies may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. Results obtained using any such assay on a sample

from a plant transformant or a progeny thereof is compared with those from a non-transformed source as a control.

The eukaryotic host cells of this invention are grown under favorable conditions to effect transcription of the transgene. Non-limiting examples of eukaryotic hosts are fungus, yeast, plant cells, insect, avian, mammalian or other animal cells. The host cells can be used, *inter alia*, as repositories of the transgene and/or vehicles for production of the transgene-specific double stranded RNAs. The host cells may also be employed to generate transgenic organisms such as transgenic animals and plants comprising the recombinant DNA vectors of the present invention. Preferred host cells are those having the propensity to regenerate into tissue or a whole organisms. Examples of these preferred host cells are oocytes, blastocutes, and certain plant cells exemplified herein.

Accordingly, this invention provides transgenic plants carrying the subject vectors. In a preferred embodiment, the transgenic plant exhibits a reduced expression (when compared to a control plant) of an endogenous gene that is substantially homologous to the transgene carried in the subject vector.

The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, Methods for Plant Molecular Biology, Mary A. Shuler and Raymond E. Zielinski, Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil.

The regeneration of plants containing the subject vector introduced by Agrobacterium tumefaciens from leaf explants can be achieved as described by Horsch et al., *Science*, 227:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley et al., *Proc. Natl. Acad. Sci. U.S.A.*, 80:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transformant shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil to allow

the production of roots. These procedures will vary depending upon the particular plant species employed, as is apparent to one of ordinary skill in the art.

A population of progeny can be produced from the first and second transformants of a plant species by methods well known in the art including cross fertilization and asexual reproduction. Transgenic plants embodied in the present invention are useful for production of desired proteins, and as test systems for analysis of the biological functions of a gene.

**Uses of the vectors of the present invention:**

The subject vectors provide specific reagents for inhibiting expression of an endogenous gene present in a host cell. The expression inhibition methods may be used in a wide variety of circumstances including suppression of a gene associated with a particular disease or disease stage; delineating the biological functions of a gene by analyzing a phenotypic change in the host cell that correlates with the selective suppression of gene expression; and facilitating drug screening by rendering the host cell more susceptible or resistant to a therapeutic agent of interest.

Accordingly, this invention provides a method of inhibiting expression of an endogenous gene present in a eukaryotic cell. The method comprises the steps of: (a) providing a subject vector containing a transgene that is substantially homologous to an endogenous gene of a eukaryotic cell; (b) introducing the recombinant vector into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

In a separate embodiment, the invention provides a method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene. The method involves: (a) providing a recombinant vector of the present invention, wherein the transgene contained in the vector is substantially homologous to the endogenous gene; (b) introducing the recombinant vector of (a) into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and

(d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell.

The host cells encompassed by these embodiments are eukaryotic cells susceptible to dsRNA-mediated "genetic interference". dsRNA induced gene silencing has been observed in a variety of multi-cellular organisms including but not limited to worms, fruitflies, protozoa, fungi, mammals, and zebrafish. Thus, cells from any of these exemplary organisms can be employed. Suitable host cells may be derived from primary cultures or subcultures generated by expansion and/or cloning of primary cultures. Any cells capable of growth in culture can be used as host cells. Of particular interest is the type of cell that differentially expresses (over-expresses or under-expresses) a disease-causing gene. As is apparent to one skilled in the art, various cell lines may be obtained from public or private repositories. The largest depository agent is American Type Culture Collection (<http://www.atcc.org>), which offers a diverse collection of well-characterized cell lines derived from a vast number of organisms and tissue samples.

Upon delivery of the subject vectors, the host cells are cultured under conditions favorable for gene transcription. The parameters governing eukaryotic cell survival are generally applicable for induction of gene transcription. The culture conditions are well established in the art. Physicochemical parameters which may be controlled *in vitro* are, e.g., pH, CO<sub>2</sub>, temperature, and osmolarity. The nutritional requirements of cells are usually provided in standard media formulations developed to provide an optimal environment. Nutrients can be divided into several categories: amino acids and their derivatives, carbohydrates, sugars, fatty acids, complex lipids, nucleic acid derivatives and vitamins. Apart from nutrients for maintaining cell metabolism, most cells also require one or more hormones from at least one of the following groups: steroids, prostaglandins, growth factors, pituitary hormones, and peptide hormones to survive or proliferate (Sato, G.H., et al. in "Growth of Cells in Hormonally Defined Media", Cold Spring Harbor Press, N.Y., 1982; Barnes and Sato (1980) *Anal. Biochem.*, **102**:255. Given the vast wealth of information on the nutrient requirements, medium conditions optimized for cell survival, one skilled in the art can readily fashion various culture conditions using

any one of the aforementioned methods and compositions, alone or in any combination.

The inhibition of expression of the endogenous gene sharing substantial sequence homology with the transgene carried in the vectors can be determined by assaying for a difference, between the host cell and the control cell, in the level of mRNA transcripts of the endogenous gene. Alternatively, a suppression in expression is determined by detecting a difference in the level of the polypeptide(s) encoded by the endogenous gene. A preferred method is to detect a phenotypic change resulting from the decrease in expression of the endogenous gene of interest.

In assaying for an alteration in mRNA level, nucleic acid contained in the host cells is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), *supra* or extracted by nucleic-acid-binding resins following the accompanying instructions provided by manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by hybridization (e.g. Northern blot analysis) and/or amplification procedures according to methods widely known in the art or based on the methods exemplified herein.

Reduction in expression of the endogenous gene can also be determined by examining the protein product of the endogenous gene. A variety of techniques is available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, *in situ* immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE. In addition, cell sorting analysis can be employed to detect cell surface antigens. Such analysis involves labeling target cells with antibodies coupled to a detectable agent, and then separating the labeled cells from the unlabeled ones in a cell sorter. A sophisticated cell separation method is fluorescence-activated cell sorting (FACS). Cells traveling in single file in a fine stream are passed through a laser beam, and the fluorescence of each cell bound by the fluorescently labeled antibodies is then measured.

Antibodies that specifically recognize and bind to the protein products of interest are required for conducting the aforementioned protein analyses. These antibodies may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra*, and 5 Sambrook et al. (1989) *supra*.

Inhibition of gene expression can also result in phenotypic change(s) in a host cell. As used herein, phenotypic change refers to any non-genotypic change that can be detected visually, or analyzed biochemically or genetically. The choice of detection methods will largely depend on the nature of the phenotypic 10 characteristics that are under investigation. For instance, certain phenotypic features of a plant cell can be detected microscopically or macroscopically. These features include improved tolerance to herbicides, improved tolerance to extremes of heat or cold, drought, salinity or osmotic stress; improved resistance to pests (insects, nematodes or arachnids) or diseases (fungal, bacterial or viral), production of 15 enzymes or secondary metabolites; male or female sterility; dwarfness; early maturity; improved yield, vigor, heterosis, nutritional qualities, flavor or processing properties, and the like. Other detectable phenotypic changes are morphological alterations including but not limited to stunting, hyperbranching, vein banding, ring spot, etching, and those responsible for color characteristics including bleaching and 20 chlorosis.

For animal cells, detectable phenotypic changes may encompass alterations in cell cycle regulation, cell differentiation, apoptosis, chemotaxis, cell motility and cytoskeletal rearrangement. Methods for detecting these phenotypic changes are well-established in the art and hence are not detailed herein.

Other phenotypic changes commonly observed in both plant and animal cells 25 involve differential expression (over-expression or under-expression) of a particular protein due to the selective inhibition of the endogenous gene of interest. Differential gene expression may be analyzed by any chemical means available in the art or those disclosed herein. As is also apparent to artisans, altering expression 30 of one endogenous gene may lead to changes in gene expression profile of a host of genes mapped to the same or related signal transduction pathways. As used herein, "signal transduction" refers to the process by which stimulatory or inhibitory signals are transmitted into and within a cell to elicit an intracellular response. Any

fluctuation in intracellular response of a eukaryotic host cell is also considered as a type of phenotypic change.

Alteration in intracellular response is often determined with the aid of reporter molecules. For example, when examining a signaling cascade involving a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the 5 signaling pathway of a trimeric G<sub>q</sub> protein is analyzed, calcium-sensitive fluorescent probes can be employed as reporters. As is apparent to artisans in the field of signal transduction, trimeric G<sub>q</sub> protein is involved in a classic signaling pathway, in which activation of G<sub>q</sub> stimulates hydrolysis of phosphoinositides by phospholipase C to 10 generate two classes of well-characterized second messengers, namely, diacylglycerol and inositol phosphates. The latter stimulates the mobilization of calcium from intracellular stores, and thus resulting in a transient surge of 15 intracellular calcium concentration, which is a readout measurable with a calcium-sensitive probe.

Another exemplary class of reporter molecules is a reporter gene operably linked to an inducible promoter that can be activated upon the stimulation or inhibition of a signaling pathway. Reporter proteins can also be linked with other proteins whose expression is dependent upon the stimulation or suppression of a given signaling cascade. Commonly employed reporter proteins can be easily 20 detected by a colorimetric or fluorescent assay. Non-limiting examples of such reporter proteins include : β-galactosidase, β-lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Those skilled in the art will know of other suitable reporter molecules 25 for assaying changes in a specific signaling transduction readout, or will be able to ascertain such, using routine experimentation.

To discern inhibition of gene expression, one typically conducts a comparative analysis of the subject and appropriate controls. Preferably, a test includes a positive control sample exhibiting a decrease in gene expression and a negative control having an unaltered expression level. The selection of an 30 appropriate control cell or tissue is dependent on the sample cell or tissue initially selected and its phenotype which is under investigation.

In one aspect, the invention methods can be employed to selectively inhibit expression of an endogenous gene that is native to the eukaryotic host cell. Such a gene may encode encodes a protein selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein and a chaperon protein. Of particular interests are endogenous genes that confer phenotypic changes as a result of inhibition of the expression and/or function of the endogenous genes. In another aspect within this embodiment, the endogenous gene is heterologous to the host cell. As used herein, heterologous genes are acquired exogenously by the host cell. Non-limiting examples of heterologous genes are those derived from virus, bacterium, fungus, and protozoa.

In a separate embodiment, the invention methods are used to identify a biological function(s) of an endogenous gene in a eukaryotic cell by examining a phenotypic change associated with the inhibition in its expression and thus loss of biological function. In essence, the subject methods allow the creation of a transient or more long-term gene-specific knock-out system for analyzing the biological function of any endogenous gene of interest.

#### Kits comprising the vectors of the present invention

The present invention also encompasses kits containing the vectors of this invention in suitable packaging. Kits embodied by this invention include those that allow generation of a double-stranded RNA transcript in a eukaryotic cell.

Each kit necessarily comprises the reagents which render the delivery of vectors into a eukaryotic host cell possible. The selection of reagents that facilitate delivery of the vectors may vary depending on the particular transfection or infection method used. The kits may also contain reagents useful for generating labeled polynucleotide probes or proteinaceous probes for detection of gene silencing. Each reagent can be supplied in a solid form or dissolved/suspended in a liquid buffer suitable for inventory storage, and later for exchange or addition into the reaction medium when the experiment is performed. Suitable packaging is provided. The kit can optionally provide additional components that are useful in the procedure. These optional components include, but are not limited to, buffers, capture reagents, developing reagents, labels, reacting surfaces, means for detection, control samples, instructions, and interpretive information. The kits can be

employed to generate eukaryotic cells whose endogenous genes are selectively inhibited, and transgenic organisms comprising these eukaryotic cells.

Further illustration of the development and use of vectors and assays according to this invention are provided in the Example section below. The examples are provided as a guide to a practitioner of ordinary skill in the art, and are not meant to be limiting in any way.  
5

## EXAMPLES

Example 1: Construction of recombinant vectors comprising two opposing transcription units

5

We have designed a recombinant vector construct useful for silencing nuclear genes in many of the agriculturally-important cereal crops. The vector comprises sequences derived from maize streak geminivirus, isolated MSV-Kom (genbank accession number AF003952, classification: Family *Geminiviridae*, genus *Mastrevirus*, species maize streak virus, designated MSV-Komatipoort. Maize streak virus has a broad host range that encompasses all agriculturally important cereal crops, including but not limited to corn, wheat, rice, barley, rye, sorghum and millet. The methods for construction of infectious geminiviruses are well known to those skilled in the art, and are described in European patent application 8687015.5 as well as in US Patent No. 5,569,597.

10

15

20

25

We have synthesized a 1618 base pair synthetic DNA that contains the MSV-Kom *repA* and *repB*, long intergenic region (LIR) and short intergenic region (SIR) and thus all sequences that are required for viral replication. Palmer et al.(1999) *Archives of Virology* 144:1345-1360. This fragment was cloned into the pZeRO-2 vector (Invitrogen) as an *EcoRI-XbaI* fragment, to create the plasmid pMSVLSB-1, the sequence of which is shown in Figure 4. A 171 base pair fragment containing the movement protein (mp) promoter of MSV-Kom is synthesised and cloned into the pZeRO-2 vector as an *HindIII-EcoRI* fragment to create pMSVLSB-2 (sequence shown in Figure 5). The *ApaI* fragment containing the mp promoter is inserted between the two *ApaI* sites in pMSVLSB-1, to create pMSVLSB-3 (sequence shown in Figure 6).

30

The cauliflower mosaic virus 35S RNA promoter (CaMV 35S promoter) sequence is amplified with a vector containing this sequence (pBI121, from Clontech) as template DNA, using the following PCR primers containing the following restriction sites (shown in italicized): *EcoRI* in CaMV35SF and *SaII* in CaMV35SR.

CaMV35SF:

TTTGAATTCGTCAACATGGTGGAGCAC (SEQ ID NO:1)

CaMV35SR:

TTTGTGACGTCCCTCTCAAATGAAATGAAC (SEQ ID NO:2)

5

The CaMV 35S promoter PCR product yielded is digested with *Eco*RI and *Sa*II and the restricted fragments are purified.

10 The zeocin resistance gene is amplified by PCR with the vector pZeRO-1 (Invitrogen) as template, using the following primers containing the following restriction sites shown in italicized: *Sa*II, *Pac*I and *Not*I in ZeoF and *Xho*I, *Pac*I and *Not*I in ZeoR:

ZeoF:

15 CCCGTCGACTTAATTAAGCGGCCGCGTTACAATTTCGCCTGATGC  
(SEQ ID NO:3)

ZeoR:

20 CCCCTCGAGTTAATTAAGCGGCCGCTCAAAAGGATCTCACCTA  
G (SEQ ID NO:4)

The zeocin resistance gene product yielded is digested with *Xho*I and *Sa*II and purified.

25 The nopaline synthase (nos) terminator sequence is amplified by PCR with the vector pBI121 (Clontech) as template, using the following primers, with restriction sites *Xho*I in nosF and *Spe*I in nosR italicized:

NosF:

30 TTTCTCGAGCGAATTCCCCGATCGTTAACAC (SEQ ID NO:5)

NosR:

TTTACTAGTCCCGATCTAGAACATAGATGAC (SEQ ID NO:6)

The nos terminator product yielded is digested with *Xba*I and *Spe*I and purified.

5        The digested CaMV35S promoter, zeocin resistance gene and nos terminator sequences are ligated together with T4 DNA ligase. The ligated product is diluted 1:100 in sterile water and the whole ligation product is re-amplified with the CaMV35SF and nosR primers. The resulting PCR product is digested with *Eco*RI and *Spe*I, purified and ligated with pMSVLSB-3 that is pre-digested with *Eco*RI and *Spe*I. The ligation reaction is used to transform *E. coli* competent cells.  
10      Transformants are selected on Luria Agar plates containing both kanamycin (100 µg/ml) and zeocin (50 µg/ml) to select for colonies containing the CaMV35S promoter-zeocin resistance gene-nos terminator cassette inserted into pMSVLSB-3 (Figure 6 and SEQ ID NO:11). Colonies putatively containing the correct plasmid  
15      are chosen, plasmid DNA isolated and screened by digestion with *Eco*RI and *Spe*I. One plasmid designated pMSVLSB-4 (Figure 7 and SEQ ID NO:12) is selected.

One of the methods in the art of construction of infectious clones of geminivirus genomes is to clone tandemly duplicated sequences of the geminivirus genome, with at least the LIR duplicated. This allows the virus sequence to escape from the cloning vector *in planta* by a replicative release mechanism. The virus Rep protein is transiently expressed in transfected cells, and induces a nick at each of the stem loop sequences contained within the origin of replication in the LIR. Rolling circle replication is initiated at each nick point, and this results in release of a ssDNA copy of the virus replicon, which is circularized by the Rep protein, and which then replicates autonomously in the plant cell nucleus. The *Xba*I-*Spe*I fragment from pMSVLSB-3, containing the viral LIR and Rep genes is inserted into the unique *Spe*I site in pMSVLSB-4 to create pMSVLSB-5 (Figure 8 and SEQ ID NO:13). The zeocin resistance gene is deleted by digestion with *Not*I; the DNA is recircularized and used to transform *E. coli* to kanamycin resistance with a new vector, pMSVLSB-6 (Figure 9 and SEQ ID NO:14). When the vector is introduced into plant cells, a monomeric copy of the insert is released by replicative release (described above) and replicates autonomously as construct MSVLSB-6 in the nuclei of infected cells.

The restriction map of construct MSVLSB-6 is shown in Figure 3; this genetic construct possesses the following features: (a) the *rep* genes and origins of replication from maize streak geminivirus that are necessary and sufficient for the autonomous replication of the viral construct and its associated foreign DNA in the host plant cell; (b) two overlapping transcription units present in the DNA replicon. The two overlapping transcription units are arranged according to the configuration shown in Figure 2. With reference to Figure 2, "promoter 1" and "terminator 1" in MSVLSB-6 are the MSV mp promoter and transcription termination signals present in the SIR, respectively, and "promoter 2" and "terminator 2" are the CaMV 35S RNA promoter and nos terminator sequences, respectively. The two overlapping transcription units share three unique restriction sites (*Sall*, *PacI* and *NotI*) and one non-unique restriction site (*XbaI*) where foreign DNA may be inserted so that it may be transcribed by both promoters to yield at least a partially double stranded RNA duplex of the foreign DNA sequence.

15

Example 2: Use of recombinant vectors to inhibit or silence gene expression in cereal crops:

20

*Application of pMSVLSB-6 in inhibition of Dwarfl gene expression in rice*

25

The vector pMSVLSB-6 exemplified above can be employed to inhibit expression of any endogenous gene in a variety of plant host cells. By way of illustration, the rice gene *Dwarf1* is inhibited to duplicate known mutant phenotype using a pMSVLSB-6 containing a fragment of the coding sequence of *Dwarf1* (Genbank accession number AB028602). The gene is amplified from cDNA isolated from rice seedlings. Primer sequences are designed to have homology with the published sequence of *Dwarf1*. Ashikari *et al.* (1999) *PNAS U.S.A.* **96**:10284-10289. The primer sequences contain *NotI* restriction sites at their 5' ends. The PCR product is digested with *NotI* and cloned into the *NotI* site of pMSVLSB-6 to generate pMSVLSB-6::dwarf1s and pMSVLSB-6::dwarf1a, with the insert cloned in the sense and antisense orientation with respect to the MSV mp promoter, respectively. The *XbaI-SpeI* fragment from each of these plasmids is transferred into an *Agrobacterium* binary vector that is commonly used for rice transformation. This vector is used to transform electrocompetent *Agrobacterium* strain LBA4404

(Life Technologies). *Agrobacterium* cultures containing the appropriate plasmids are used in transformation of rice. Transgenic rice is generated by standard protocols (see, e.g. US Patent 5,591,616). The transgenic rice plants display similar phenotypes to the *dwarf1* mutant described by Ashikari *et al.* (1999) *supra*: they are giberellin-insensitive, dwarfed in comparison with un-silenced transgenic controls, and having broad, dark green leaves, compact panicles and short, round grains.

5  
10  
*Application of pMSVLSB-6 in inhibition of phytoene desaturase expression in maize seedlings*

15  
The coding sequence for the maize phytoene desaturase gene (*pds*), having the Genbank accession number U37285, is amplified from cDNA made from RNA isolated from four-day-old maize seedlings, of the cultivar "Golden Cross Bantam". The primers used for amplification of this cDNA have the following sequences containing the *PacI* sites (italicized) at the 5' ends:

20  
zeapds1330:  
TTTTTA~~ATTA~~AAGGTCCGCCTGAATTCTCG (SEQ ID NO:7)

25  
zeapds1873  
TTTTTA~~ATTA~~ACGGCAAGGCTCACAGTTG (SEQ ID NO:8)

PCR amplification with these primers and cDNA made from RNA isolated from maize seedlings yields a product of 565 base pairs, which is then digested with *PacI*. The progenitor plasmid to pMSVLSB-6, pMSVLSB-5 is digested with *XbaI* and *SpeI* to release the MSV and associated overlapping transcription unit sequences from the pZeRO-2 cloning vector as a single 4816 base pair fragment. This fragment is cloned into the *Agrobacterium* binary vector pBin19 (Genbank: U09365) digested with *XbaI* to yield pMSVLSB-7. The plasmid pMSVLSB-7 is digested with *PacI* and the *pds* PCR fragment is inserted into this position, generating plasmid pMSVLSB-7::*pds1* (cloned in the sense orientation with respect to the MSV mp promoter) and pMSVLSB-7::*pds2* (cloned in the antisense orientation with respect to the MSV mp promoter). These two plasmids are each

introduced into *Agrobacterium* strain C58C1(pMP90) (Koncz and Schell, 1985) by electroporation. The *Agrobacterium* containing the binary vector plasmids is grown overnight in Luria Bertani medium containing appropriate selective antibiotics. The bacterial suspension is loaded into a 100 µl Hamilton syringe and injected into three day old maize seedlings (cultivar Golden Cross Bantam) according to methods described by Escudero et al. (1994) in the chapter "Agroinfection" of The Maize Handbook, Freeling M, Walbot V (eds). Plants that are successfully agroinfected display a photobleaching phenotype on the first three leaves, similar to that induced by spraying the plants with the phytoene desaturase-inhibitor norfluorazon.

10

CLAIMS

What is claimed is:

1. A eukaryotic recombinant vector comprising a viral replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene in a eukaryotic host cell.
2. The eukaryotic recombinant vector of claim 1, wherein each of the overlapping transcription units comprises a promoter and a terminator.
3. The eukaryotic recombinant vector of claim 2, wherein the promoter is a constitutive promoter.
4. The eukaryotic recombinant vector of claim 2, wherein the promoter is an inducible promoter.
5. The eukaryotic recombinant vector of claim 2, wherein the promoter is a tissue-specific promoter.
6. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(a).
7. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(b).
8. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(c).

9. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(d).

5 10. The eukaryotic recombinant vector of claim 1 that inhibits gene expression of the eukaryotic host cell.

10 11. The eukaryotic recombinant vector of claim 1, wherein the eukaryotic host cell is selected from the group consisting of fungus, yeast cell, plant cell and animal cell.

15 12. The eukaryotic recombinant vector of claim 1 that inhibits expression of an endogenous gene present in the host cell, wherein the endogenous gene is substantially homologous to the transgene contained in the overlapping transcription units.

13. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is native to the host cell.

20 14. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is heterologous to the host cell.

25 15. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa.

16. The eukaryotic recombinant vector of claim 1, wherein expression of the transgene to yield double-stranded RNA transcripts confers a phenotypic change in the eukaryotic host cell.

30 17. The eukaryotic recombinant vector of claim 1, wherein the transgene encodes a protein selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, and a chaperon protein.

18. The eukaryotic recombinant vector of claim 1 that is an autonomously replicating vector.

5 19. The eukaryotic recombinant vector of claim 1, wherein the viral replicon is derived from a DNA virus.

10 20. The eukaryotic recombinant vector of claim 19, wherein the DNA virus is selected from the group consisting of *Geminivirus*, *Caulimoviridae*, *Badnaviridae*; *Circoviridae*, *Circinoviridae*, *Parvoviridae*, *Papovaviridae*, *Polyomaviridae*, *Adenoviridae*, *Herpesviridae*, *Poxviridae*, *Iridoviridae*, *Baculoviridae*, *Hepadnaviridae*, *Retroviridae*, *Gyrovirus*, *Nanovirus*, and African Swine Fever virus.

15 21. A host cell transformed with a vector of claim 1 or 10.

20 22. The host cell of claim 21 that is a eukaryotic cell selected from the group consisting of fungus, yeast cell, plant cell and animal cell.

25 23. A transgenic plant comprising a eukaryotic recombinant vector of claim 1 or 10.

24. The transgenic plant of claim 23 exhibiting reduced expression of an endogenous gene that is substantially homologous to the transgene contained in the eukaryotic recombinant vector.

30 25. A kit for generating a double-stranded RNA transcript in a eukaryotic cell comprising a eukaryotic recombinant vector of claim 1 in suitable packaging.

26. A method of inhibiting expression of an endogenous gene present in a eukaryotic cell, comprising:

(a) providing a eukaryotic recombinant vector of claim 12;

- (b) introducing the eukaryotic recombinant vector into the eukaryotic cell;
- (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene that is contained in the transcription units of the vector, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

5

27. The method of claim 26, wherein the endogenous gene is native to the host cell.

10

28. The method of claim 26, wherein the endogenous gene is heterologous to the host cell.

15

29. The method of claim 26, wherein the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa.

20

30. The method of claim 26, wherein inhibition of the endogenous gene confers a phenotypic change in the host cell.

31. The method of claim 26, wherein the host eukaryotic cell is selected from the group consisting of fungus, yeast cell, plant cell, and animal cell.

25

32. The method of claim 26, wherein the eukaryotic recombinant vector is an autonomously replicating vector.

33. The method of claim 26, wherein the eukaryotic recombinant vector comprises a viral replicon derived from a DNA virus.

30

34. The method of claim 26, wherein the DNA virus is selected from the group consisting of *Geminivirus*, *Caulimoviridae*, *Badnaviridae*; *Circoviridae*, *Circinoviridae*, *Parvoviridae*, *Papovaviridae*, *Polyomaviridae*, *Adenoviridae*,

*Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retrovirida, Gyrovirus, Nanovirus, and African Swine Fever virus.*

35. The method of claim 26, wherein the eukaryotic recombinant vector  
5 comprises two overlapping transcription units, wherein each transcription unit  
comprises a promoter and a terminator.

36. The method of claim 26, wherein the promoter is a constitutive promoter.  
10 37. The method of claim 26, wherein the promoter is an inducible promoter.

38. The method of claim 26, wherein the promoter is a tissue-specific  
promoter.

15 39. The method of claim 35, wherein the promoter and the terminator of the  
overlapping transcription units are arranged in a configuration shown in Figure 2(a).

40. The method of claim 35, wherein the promoter and the terminator of the  
overlapping transcription units are arranged in a configuration shown in Figure 2(b).  
20

41. The method of claim 35, wherein the promoter and the terminator of the  
overlapping transcription units are arranged in a configuration shown in Figure 2(c).

25 42. The method of claim 35, wherein the promoter and the terminator of the  
overlapping transcription units are arranged in a configuration shown in Figure 2(d).

43. A method of identifying a biological function(s) of an endogenous gene  
of interest in a eukaryotic cell by selectively inhibiting the expression of the  
endogenous gene, the method comprising:

- 30 (a) providing a eukaryotic recombinant vector of claim 12;  
(b) introducing the eukaryotic recombinant vector of (a) in to the  
eukaryotic cell;

- (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the eukaryotic recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell;
- 5 and
- (d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell.

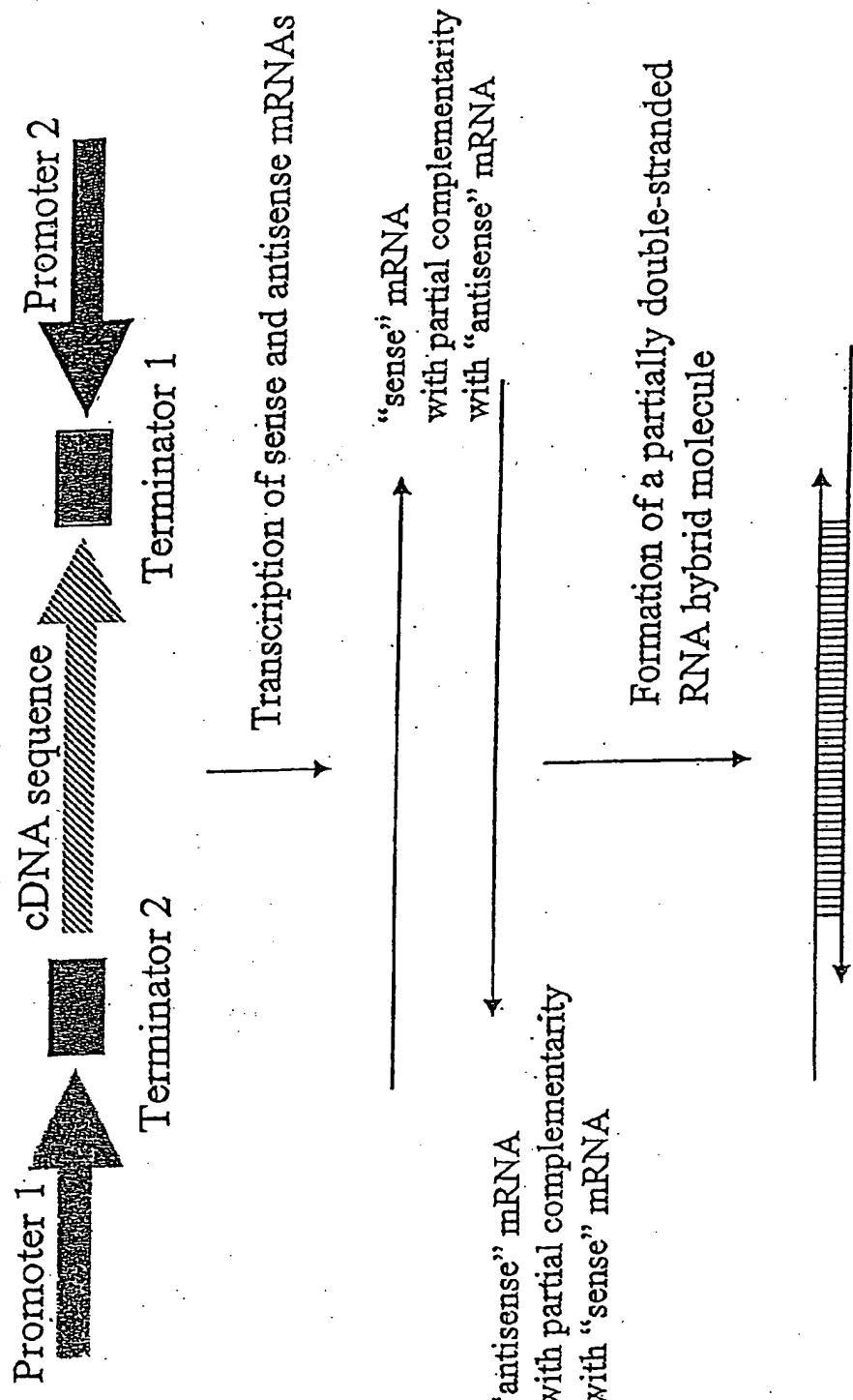
10

44. The method of claim 43, wherein the eukaryotic cell is selected from the group consisting of fungus, yeast cell, plant cell, and animal cell.

45. The method of claim 43, wherein the eukaryotic cell is a plant cell.

15

46. The method of claim 43, wherein the eukaryotic cell is an animal cell.

**Figure 1**

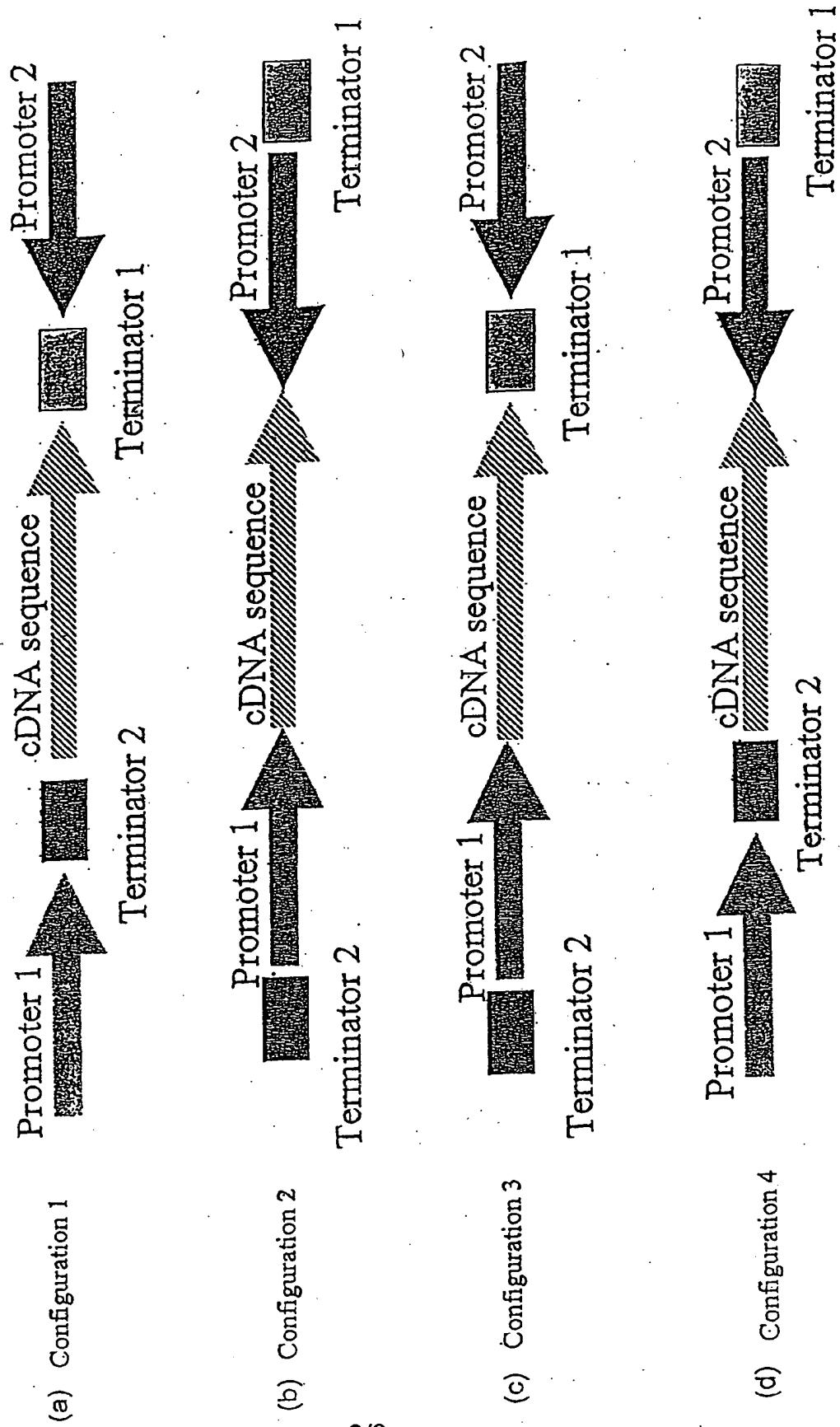
**Figure 2**

Figure 3

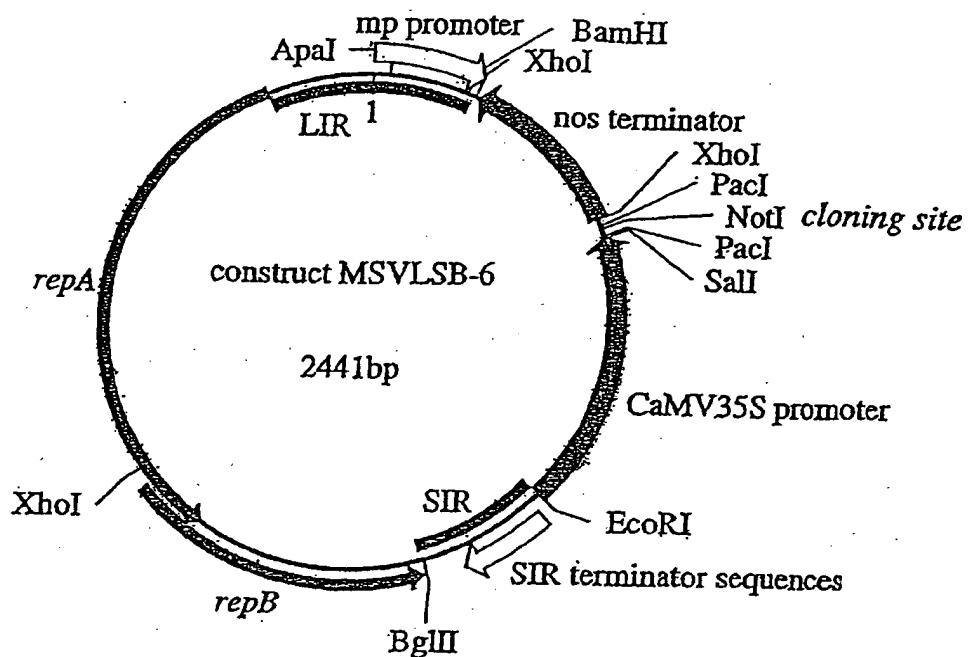


Figure 4

pMSVLSB-1: 4881 bp;  
 Composition 1161 A; 1260 C; 1251 G; 1209 T; 0 OTHER  
 Percentage: 24% A; 26% C; 26% G; 25% T; 0% OTHER

Molecular Weight (kDa) ssDNA: 1506.65 dsDNA: 3009.2

## ORIGIN

```

1      AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC
61     ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC
121    TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181    TTGTGAGCGG ATAACAATT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241    TTAGGTGACA CTATAGAATA CTCAAAGCTAT GCATCAAGCT TGTTACCGAG CTCGGATCCA
301    CTAGTAACGG CGGCCAGTGT GCTGGAATTIC ATGGGCAGAC CGCTCTGTAC TTTAAGAGTG
361    TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTGTGATGAAT GCTGAAAAGCT
421    TACAAITAATA TGTCGTGCGA TGGCACGAAA AAACACACCC AAACAAATACA GGGGGGTAGT
481    CGGGGGGGCGG CTAAAGGGTGG TGCTCGGCGG CGAACACATC GAAAAATCAA GATCTATATG
541    AATTACACTT CCTCCGTAGG AGGAACCCAA GGGGGAGAAAT ACCACTCTC CCCCAGGCGAC
601    ATAATGTAAA TGACCGAGTT TGCCTCGAAA TACTCTCAGCT GCCCTGGAGT CATTTCCTTC
661    ATCCAATCTT CATCCGAGTT GGCAGGAGT ATTGTAGGGCT TAGACTTCTT CTGCACCTTT
721    TCTCTCTTAC CATACTTGGG GTTTACATG AAATCCCTCT GACAGCCAAC TAATGTTTC
781    CAACAAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTIG
841    TATGAAGACC AATCAACATT ATTTTGCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA
901    GTAGATTTTC CGGTTCTGT TGGGGCGACG ATGTAGAGGC TCTGCTTCT TGATCTTCA
961    TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAAG AATTGCAATCC TCGAGGGTAT
1021   AACAGGTAGG TTGAAGGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAAGATG TTAGGCTGGA
1081   GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGG TGAGGATTGG
1141   TGAACCTTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT
1201   TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATEGGA GAGGTACTCT TCTTGGAGG
1261   TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTITAGA AGGCTTTTC TCCTTACCT
1321   CTGAATCAGA TTTTCTAGG AAGGGGGACT TCCTAGGART GAARGTACCT CTCTCAAACA
1381   CAGCCAGAGG TTCCCTGAGA ATGTAATCCC TCACTCTGT AACTGACTTIG GCACCTGAA
1441   TATTTGGGTG AAACCCATT ATATCAAAGA ACCTTIGAGTC AGATATCCTT ATCGCTTCT
1501   CTGGCTGTGA CRATGCTTGT AAATGCAAC TTCCATCTT ATGTCCTCT CGGGCACATA
1561   GAATATATTG GGGAAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG CGGATTTCAAG
1621   GATTTCTGG ACACTTGGG TAGTTTAGGA ACCTGTTAGC GTTCTCTGTG GAGAACTGAC
1681   GGTGGGAATGA GGAGGGAGCC ATAGCCGACG ACGGAGGGTG AGGCTGAGGG ATGGCAGACT
1741   GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCCTTCGAA ATCCGCCGCT CCATTGTCTT
1801   ATAGTGGTTC TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA
1861   TATTACCGG CCGTCTTTC CTGCGAGGGC CGGTTAGGGA CGAGCGCCT TGATTTAAAG
1921   CCTGGTTCTG CTTTGGGGCC GCTCGAGCAT GCATCTAGAG GGCCCAATTC GCCCTATAGT
1981   GAGTCGTATT ACAATTCACT GGCCGTGCGT TTACAACGTC GTGACTGGGA AAACCCGTGGC
2041   GTTACCCAACT TAAATCGCCT TGCAGCACAT CCCCTTTCG CCAGCTGGCG TAAATAGCGAA
2101   GAGGGCCGCA CGGATCGCCC TTCCCAACAG TTGCGCAGCC TATACTGACG GCAGTTAAAG
2161   GTTTACACCT ATAAAAGAGA GAGCCGTTAT CGTCTGTGTTG TGGATGTACA GAGTGATATT
2221   ATTGACACGC CGGGGCGACG GATGGTGATC CCCCTGGCCA GTGCACGTCT GCTGTCAGAT
2281   AAAGTCTCCC GTGAACCTTA CCCGGTGGTG CATATGGGGG ATGAANCTG CGGCATGATG
2341   ACCACCGATA TGGCCAGTGT CGGGGTCTCT GTTATCGGGG AAGAAGTGGC TGATCTCAGC
2401   CACCCCGAAA ATGACATCAA AAACGCCATT AACCTGATGT TCTGGGAAAT ATAAATGTCA
2461   GGCTGAAATG GCGAATGGAC GCGCCCTGTA CGGGCGCATT AAGCGCGCGG GTGTGGTGGT
2521   TACCGCGAGC GTGACCGCTA CACTTGGCAG CGCCCTAGCG CCCGCTCCTT TCGCTTCTT
2581   CCCCTCTTCTT CTGCCCCAGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC
2641   TTTAGGGTTC CGATTTAGAG CTTTACGGCA CCTCGACCGC AAAAAACTTG ATTTGGGTGA
2701   TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA CGTTGGAGTC
2761   CACGTTCTTT AATAGTGGAC TCTTGTCTCA AACTGGAACA ACACCAACCTT CTATCGCGGT
2821   CTATTCTTTT GATTTATAAG GGATGTGCCC GATTTGGGCC TATGGTTAA AAAATGAGCT
2881   GATTTAACAA AAATTTAAC AAAATTCAAG AGAACCTCGTC AAGAAGGGCA TAGAAGGGCA

```

Figure 4 (cont'd)

2941 TCGCCTGCGA ATCGGGAGCG GCGATACCGT AAAGCACGAG GAAGCGGTCA GCCCATTGCG  
 3001 CGCCAAGCTC TTCAGCARATA TCACGGTAG CCAACGCTAT GTCCCTGATAG CGGTCCGCCA  
 3061 CACCCAGCCG GCCACAGTCG ATGARNTCCAG AAAAGCGGCC ATTTCACC ATGATATTG  
 3121 GCAAGCAGGC ATCGCCATGG GTCACGACGA GATCCCTGCC GTCGGGCATG CTGGCCTTG  
 3181 GCCTGGCAA CAGTTGGGT GGCGCGAGCC CCTGATGCTC TTGCTCCAGA TCATCCTGAT  
 3241 CGAACAGGACG GGCTTCCATC CGAGTACGTG CTCGCTCGAT GCGATGTTTC GCTTGGTGGT  
 3301 CGAATGGCA GGTAGCGGGA TCAAGCGTAT GCAGCCGCCG CATTGCATCA GCCATGATGG  
 3361 ATACTTTCTC GGCAAGGAGCA AGGTGAGATG ACAGGAGATC CTGCCCCGGC ACTTCGCCCCA  
 3421 ATAGCAGCCA GTCCCCTTCCC GCTTCAGTGA CAACGTCGAG CACAGCTGCG CAAGGAACGC  
 3481 CGCTCGTGGC CAGCCACGAT AGCCGCGCTG CCTCGTCTTG CAGTTCATTC AGGGCACCGG  
 3541 ACAGGTGGT CTTGACAAA AGAACCGGGC GCCCCTGCCG TGACAGCCGG AACACGGCGG  
 3601 CTCAGAGCA GCCGATTGTC TGTTGTGCC AGTCATAGCC GAATAGCCTC TCCACCCAAG  
 3661 CGGGCGGAGA ACCTGCGTGC AATCCATCTT GTTCAATCAT GCGAAACGAT CCTCATCTG  
 3721 TCTCTTGATC AGATCTTGAT CCCCTGCCG ATCAGATCTT TGGGGCGAG AAAGCCATCC  
 3781 AGTTTACTTT GCAGGGCTTC CCAACCTTAC CAGAGGGCGC CCCAGCTGGC AATTCCGGT  
 3841 CGCTTGCTGT CCATAAAACC GCCCCAGTCTA GCTATGCCA TGTAAGCCCA CTGCAAGCTA  
 3901 CCTGCTTTCTC CTTCGGCTT GCGTTTCCC TTGTCAGAT AGCCCACTAG CTGACATTCA  
 3961 TCCGGGGTCA GCACCGTTTC TGCGGACTGG CTTTCTACGT GAAAAGGATC TAGGTGAAGA  
 4021 TCCCTTTGTA TAATCTCATG ACCAAAATCC CTTAACGTGA TTTTCGTTT CACTGAGCGT  
 4081 CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTCAGATCC TTTTTTCTG CGCGTATCT  
 4141 GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTGGCG GATCAAGAGC  
 4201 TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA AATACTGTCC  
 4261 TTCTAGTGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACCTC TGTAAGCACCG CCTACATACC  
 4321 TCGCTCTGCT AATCCCTGTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGCTTACCG  
 4381 GGTGGACTC AAGACGATAG TTACCGGATA AGCGCCAGCG GTGGGGCTGA ACGGGGGTT  
 4441 CGTGCACACA GCCCAGCTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG  
 4501 AGCTATGAGA AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG  
 4561 GCAGGGTCGG AACAGGAGAG CGCACCGAGG AGCTTCCAGG GGGAAACGCC TGGTATCTT  
 4621 ATAGTCCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTGG ATTTCCTGTA TGCTCGTCAG  
 4681 GGGGGCGGAG CCTATGGAAA AACGCCAGCA AGCGGGCCCTT TTACGGTTC CTGGGTTTT  
 4741 GCTGGCCCTT TGCCTCACATG TTCTTTCTG CGTTATCCCC TGATTCTGTG GATAACCGTA  
 4801 TTACCGCCCTT TGAGTGAAGCT GATACCGCTC GCGCAGCGG AACGACCGAG CGCAGCGAGT  
 4861 CAGTGAAGCGA GGAGCGGAA G

Figure 5

pMSVLSB-2: 3413 bp;  
 Composition 777 A; 950 C; 884 G; 802 T; 0 OTHER  
 Percentage: 23% A; 28% C; 26% G; 23% T; 0% OTHER

Molecular Weight (kDa): ssDNA: 1052.40 dsDNA: 2104.2

ORIGIN

```

1      AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGC GATTCAATTAA TGCAAGCTGGC
61     ACGACAGGT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121    TCACTCATTA GGCACCCCAG GCTTACACT TTATGCTTC GGCTCGTATG TTGTGTGGAA
181    TTGTGAGCGG ATAACAATT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241    TTAGGTGACA CTATAGAATA CTCAGCTAT GCATCAAGCT TGCGCCCGGT AGGGACCGAG
301    CGCTTGATT TAAAGCTGG TTCTGTTTG TATGATTAT CTAAGCAGC CCAATCTAAA
361    GAAACCGGTC CCGGGCACTA TAAATTGCT AACAAAGTGGC ATTCAATTCA GGATECTTTA
421    AACTCGAGTC TAGAGGGCCC GAATTGTCA GATATCCATC AACTGGCGG CGCTCGAGC
481    ATGCATCTAG AGGGCCCAAT TCGCCCTATA GTGAGTCGTA TTACAAATTCA CTGCCGTGCG
541    TTTTACAACG TCGTGACTGG GAAAACCTG CGGTACCCCA ACTTAATCGC TTGCGAGCAC
601    ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCC CGCCGATCGC CCTTCCCAAC
661    AGTTGCGCAG CCTATACTGTA CGGCAGTTA AGGTTAACAC CTATAAAAGA GAGAGCCGTT
721    ATCGTCCTGT TGTGGATGTA CGAGTGATA TTATTGACAC GCGGGGGCGA CGGATGGTGA
781    TCCCCCTGGC CAGTGCACGT CTGCTGTCAG ATAAAGTCTC CGGTGAACTT TACCCGGTGG
841    TGCAATATCGG GGATGAAAGC TGGCCATGA TGACCACCGA TATGGCCAGT GTGCCGGTCT
901    CCGTTATCGG GGAAGAAAGC GCTGATCTCA GCCACCGCGA AAATGACATC AAAACGCCA
961    TTAACCTGAT GTTCTGGGGA ATATAAAATGT CAGGCCTGAA TGGCGAATGG ACGCGCCCTG
1021   TAGCGCCGCA TTAAGCGCGC GGGTGTGGT GTACGCGCA GCGTGACCGC TACACTTGCC
1081   AGCGCCCTAG CGCCCCGCTCC TTTGCTTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCGGCG
1141   TTTCCCCGTC AAGCTCTAAA TCGGGGGCTC CTTTAGGGT TCCGATTAG AGCTTTACGG
1201   CACCTCGAACG CAAAAAAACT TGATTGGGT GTGGGTTACAC GTAGTGGGCC ATGCCCTGAA
1261   TAGACGGTTT TCGCCCTTT GACGTTGGAG TCCACGGTCT TTAAATAGTG ACTCTTGTC
1321   CAAACTGGAA CAACACTCAA CCCTATCGCG GTCTATTCTT TTGATTATA AGGGATGTTG
1381   CCGATTTCGG CCTATTGGTT AAAAATGAG CTGATTAAAC AAAAATTITA ACAAAATTCA
1441   GAAGAACTCG TCAAGAAGGC TATAGAAGGC GATAGCGCTGC GAATCGGGAG CGCGGATACC
1501   GTAAAGGACAG AGGAAGCGGT CAGCCCCATTC GCGGCCAACG TCTTCAGCAA TATCACGGGT
1561   AGCCAAACGCT ATGCTCTGAT AGCGGTCCGC CACACCCAGC CGGCCACAGT CGATGAATCC
1621   AGAAAAGCGG CCAATTTCGA CCATGATATT CGCGAACGAG GCAATGCCAT GGGTCACGAC
1681   GAGATCCTCG CCGTCGGGCA TGCTCGCTT GAGCCTGGGG AACAGTTCGG CTGGCGGAG
1741   CCCCTGATGC TCTTCGTCGA GATCATCCTG ATCGACAAAG COGGCTTCCA TCCGAGTACG
1801   TGCTCGCTCG ATGCGATGTT TCGCTGGTGT GTCGAATGGG CAGGTAGCCG GATCAAGCGT
1861   ATGCAGCCGC CGCATTGCA CAGCCATGAT GGATACCTTC TCGGCAGGGAG CAAGGTGAGA
1921   TGACAGGAGA TCCCTGCCCCG GCACCTCGCC CAATAGCAGC CAGTCCCTTC CCGCTTCAGT
1981   GACAACGTG AGCACAGTG CGCAAGGAAC GCGCGTCGTG GCCAGCCACG ATAGCAGCGC
2041   TGCCCTCGTCT TGCAGTTCAT TCAGGGCACCC GCAACAGGTCG GTCTTGACAA AAAGAACCGG
2101   GCGCCCCCTGC GCTGACAGCC GGAAACACGGC GGCATCAGAG CAGCCGATTG TCTGTGTGCG
2161   CCAGTCATAG CGGAATAGCC TCTCCACCCA AGCGGGCCGGA GAAACCTCGCT GCAATCCATC
2221   TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTTGA TCAGATCTTG ATCCCCTGCG
2281   CCATCAGATC CTGGCGCGC AGAAAGCCAT CCAGTTTACT TTGCGAGGGCT TCCCACCTT
2341   ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTCGCTTGCT GTCCATAAAA CGGCCAGTC
2401   TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTTIT CTCTTTCGCG TTGGTTTTC
2461   CCTTGTCAGG ATAGCCCCAGT AGCTGACATT CATCCGGGGT CAGCACCGTT TCTGCGGACT
2521   GGCTTCTAC GTGAAAAGGA TCTAGGTGAA GATCCCTTTT GATAATCTCA TGACCAAAAT
2581   CCCTTAACGT GAGTTTCTGT TCCACTGAGC GTCAAGACCCC GTAGAAAAGA TCAAAGGATC
2641   TTCTTGAGAT CCTTTTTTCG TGCCTGTAAT CTGCTGCTTG CAAACAAAAA AACCCACCGCT
2701   ACCAGCGGTG GTTGTGTTGC CGGATCAAGA GCTACCAACT CTTTTCCGA AGGTAACCTGG
2761   CTTCAAGAAC TCTCTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAAGTGGC
2821   CGCTGCCAGT GGCATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTACCGGA
2881

```

**Figure 5 (cont'd)**

2941 TAAGGGCCAG CGGTGGGCTT GAAACGGGGG TTTCGTGCACA CAGCCCAGCT TGGAGCGARC  
3001 GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCCA  
3061 AGGGAGAAAAG GCGGACAGGT ATCCGGTAAG CGGCAGGGTC GGAACAGGGAG AGCGCACGGAG  
3121 GGAGCTTCCA GGGGGAAACG CCTGGTATCT TTATAGTCCT GTGGGGTTC GCCACCTCTG  
3181 ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGGGG AGCCTATGGA AAAACGCCAG  
3241 CAACGCGGCC TTTTTACGGT TCCCTGGGCTT TTGCTGGCCT TTGCTGCACA TGTTCITTC  
3301 TGCCTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTGAGTGAG CTGATACCGC  
3361 TCGCCGCAGC CGAACGACCG AGGGCAGCGA GTCAGTGAGC GAGGAAGCGG AAG

## Figure 6

pMSVLSB-3:

pMSVLSB2 Apa fragment inserted: 4961 bp;  
 Composition 1190 A; 1276 C; 1262 G; 1233 T; 0 OTHER  
 Percentage: 24% A; 26% C; 25% G; 25% T; 0% OTHER

Molecular Weight. (kDa): ssDNA: 1531.26 dsDNA: 3058.5  
 ORIGIN

```

1      AGCGCCCAAT ACAGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC
61     ACGACAGGTT TCCCAGCTGG AAAGCGGGCAA GTGAGCGCAA CGCAATTAAAT GTGAGGTAGC
121    TCACTCATTA GGCACCCCGAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181    TTGTGAGCGG ATAACAATT CAACACAGGA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241    TTAGGTGACA CTATAGAATA CTCAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
301    CTAGTAACGG CCGCCAGTGT GCTGGAATTG ATGGGCAGAC CGCTCTGTC TTTAAGAGTG
361    TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTGATGAAAT GCTGAAAGCT
421    TACATTAAATA TGTCGTGCGA TGGCACGAAA AACACACAGC AAACAATACA GGGGGGTAGT
481    CGGGGGGGGG CTAAGGGTGG TGCTCGGCGG GCGAACATC GAAAAATCAA GATCTATATG
541    AATTACACIT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACITCTC CCCCAGGAC
601    ATAATGTAAA TGACGCAGTT TGCTCGAAA TACTCCAGCT GCCCTGGAGT CATTCTTC
661    ATCCAATCTT CATCCGAGTT GGCAGGATT ATTGTAGGCT TAGACTTCTT CTGCACCTT
721    TTCTCTTAC CATACTTGGG GTTACATG AAATCCCTCT GACAGCCAAAC TAATGTTTC
781    CAACAAGGAC AGAAATTTAA CGGAATATCA TCTACGATGT TGTGATTTG GTCTTCGTG
841    TATGAAGACC AATCAACATT ATTTTGCAG TAAATATGAA CCCCTAGGCT TCTGGCCAA
901    GTAGATTTTC CGGTTCTTGT TGGGCCAGC ATGTAGAGGC TCTGTTTCT TGATCTTCA
961    TCTGATGACT GGATACAGAA TCCATCATT GGAGGTCAGA AATIGCATCC TCGAGGGTAT
1021   AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGTGGA
1081   GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGG TGAGGATTGG
1141   TGAACITCTTC CTGAATCTCA GGAAAAAAGCT TATTGCGAGA GTATTCAAAA TACTGCAATT
1201   TTGTGGACCA ATCAAAGGGG AGCTCTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG
1261   TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTGA AGGCTTTTTT TCCCTTACCT
1321   CTGAATCAGA TTTCTTAGG AAGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAACAA
1381   CAGCCAGAGG TTCTTGAGA ATGTAATCCC TCACCTCTTT AACTGACTTGC GCACCTGAA
1441   TATTGGGTG AAACCCATT ATATCAAGA ACCTTGAGTC AGATATCCTT ATGGCTTCT
1501   CTGCGTGAAG CAATGCATGT AAATGCAACAC TTCCATCTT ATGTGCCTCT CGGGCACATA
1561   GAATATATTG GGGAAATCCA CGAACGACGA GCTCCCGAGAT CATCTGACAG GCGATTTCAG
1621   GATTTCCTGG ACACTTGGG TAGGTTAGGA ACGTGTAGC GTTCCCTGTT GAGAACTGAC
1681   GGTGGATGAA GGAGGAGGCC ATAGCCGACG ACGGAGGTG AGGCTGAGGG ATGGCAGACT
1741   GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCGCT CCATTGCTT
1801   ATAGTGGTGG TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA
1861   TATTACCGCG CCTTCTTTTC CTGCGAGGGC CCGGTAGGGG CCGAGCGCTT TGATTAAAG
1921   CCTGGTTCTG CTTTGTATGA TTTATCTAAA GCAGCCCAAT CTAAAGAAC CGGTCCCCGG
1981   CACTATAAT TGCCTAACAA GTGCGATTCA TTCAATGGATC CTTTAAACTC GAGTCTAGAG
2041   GGCCCAATTG GCCCTATAGT GAGTCGTATT ACAATTCACT GGCGTCTGTT TTACAACGTC
2101   GTGACTGGGA AAACCCCTGGC GTTACCCAAC TTAATGCGCT TGCAGCACAT CCCCCCTTCTG
2161   CCAGCTGGCG TAATAGCGAA GAGGCCGCA CGGATCGCCC TTCCCAACAG TTGCGCAGCC
2221   TATACTGACG CGAGTTTAAG GTTACACCT ATAAAAGAGA GAGCCCTTAT CGTCTGTTG
2281   TGGATGTACA GAGTGTATT ATTGACACGC CGGGCGCAGC GATGGTGATC CCCCCTGGCCA
2341   GTGCACTGCT GCTGTCAGAT AAAGTCCTCC GTGAACTTTA CCCCCTGGTG CATATCGGGG
2401   ATGAAAGCTG GCGCATGATG ACCACCGATA TGGCAGTGT GCGGGTCTCC GTTATCGGGG
2461   AAGAAGTGGC TGATCTCAGC CACCGGAAA ATGACATCAA AAACGCCATT AACCTGATGT
2521   TCTGGGGAAAT ATAAATGTCA GGCGTGAATG GCGAATGGAC GCGCCCTGTA GCGGCCATT
2581   AAGCGCGCGG GTGTGGTGGT TACCGCGAGC GTGACCGCTA CACTTGCAG CGCCCIAGCG
2641   CCCGCTCTT TCGCTTTCTT CCCTTCCTT CTGCGCACGT TCGCCGGCTT TCCCCGTCAA
2701   GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTAGAG CTTTACGGCA CCTCGACCGC
2761   AAAAAGCTTG ATTGGGTGA TGGTCACGT AGTGGCCAT CGCCCTGATA GACGGTTTTT

```

Figure 6 (cont'd)

2821 CGCCCTTTGA CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAAACA  
 2881 ACACTCAACC CTATCGGGT CTATTCTT GATTATAAG GGATGTTGCC GATTGGCC  
 2941 TATTGGTTAA AAAATGAGCT GATTAAACAA AAATTTAAC AAAATTCAAG AGAACCTCGTC  
 3001 AAGAAGGCAG TAGAAGGGCA TGCGCTGCGA ATCAGGGAGCG GCGATACCGT AAAGCACGAG  
 3061 GAAGCGGTCA GCCCATTGCG EGCCAAGCTC TTCAGCAATA TCACGGGTAG CCACCGCTAT  
 3121 GTTCTGATAG CGGTCCGCCA CACCCAGCG GCCACAGTCG ATGAATCCAG AAAAGCGGCC  
 3181 ATTTCCTCACC ATGATATTGCG GCAAGCAGGC ATGCCATGG GTCACGACGA GATCTCGCC  
 3241 GTCGGGCATG CTCGCTTGA GCCTGGCGAA CAGTTGGCT GGCGCAGGCC CCTGATGCTC  
 3301 TTCTGTCAGA TCATCCTGAT CGACAAGACC GGCTTCCATC CGAGTACGTG CTCGCTCGAT  
 3361 GCGATGTTTC GCTTGGTGGT CGAATGGCA GGTAAGCCGA TCAAGCGTAT GCAGCCGCCG  
 3421 CATTGATCA GCCATGATGG ATACTTCTC GGCAAGGAGCA AGGTGAGATG ACAGGAGATC  
 3481 CTGCCCCGGC ACTTCGCCA ATAGCAGCCA GTCCCTTCCC GCTTCAGTGA CAACGTCGAG  
 3541 CACAGCTGCG CAAGGAACGC CGTCTGTGGC CAGCCACGAT AGCCGCGCTG CCTGTCCTTG  
 3601 CAGTTCATTC AGGGCACCGG ACAGGTCGGT CTTGACAAAA AGAACCGGGC GCCCCCTGCGC  
 3661 TGACAGCCGG AACACGGGG CATCAGAGCA GCGGATTGTC TGTGTCGCC AGTCATAGCC  
 3721 GAATAGCCTC TCCACCCAAG CGGCCGGAGA ACCTGCGTGC AATCCATCTT GTTCATCCT  
 3781 GCGAAAAGAT CCTCATCTG TCTCTTGATC AGATCTTGTAT CCCCTGCGCC ATCAGATCTT  
 3841 TGGCGGGGAG AAAGCCATCC AGTTTACTTT GCAAGGGCTTC CCAACCTTAC CAGAGGGCC  
 3901 CCCAGCTGGC AATTCCGGT CGCTTGTGTG CCTAAACCC GCGGAGTCTA GCTATGCCA  
 3961 TGTAAGCCCA CTGCAAGCTA CCTGCTTCTC CTITGCCCTT GCGTTTCCC TTGTCAGAT  
 4021 AGCCCCAGTAG CTGACATTCA TCCGGGGTCA GCACCGTTTC TGCGGACTGG CTITCTACGT  
 4081 GAAAAGGATC TAGGTGAAGA TCCTTTTGA TAATCTCATG ACCAAAATCC CTTAACGTGA  
 4141 GTTTCTGTC CACTGAGGT CAGACCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC  
 4201 TTTTTCTG CGCGTAATCT GCTGCTTGC AAAAAAAAAA CCACCGCTAC CAGGGGTGGT  
 4261 TTGTTTGGCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC  
 4321 GCAGATACCA AATACTGTC TTCTAGTGTG GCGCTAGTTA GCGCACCACT TCAAGAACTC  
 4381 TGTAGCACCG CCTACATACC TOGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCACTGG  
 4441 CGATAAGTCG TGTCTTACCG GGTGGACTC AAGACGATAG TTACGGATA AGGCGCAGCG  
 4501 GTCGGGCTGA ACGGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA  
 4561 ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGGCCACG CTTCGGAAAG GGAGAAAGGC  
 4621 GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG  
 4681 GGGAAACGCC TGGTATCTT ATACTCTGT CGGGTTCTGC CACCTCTGAC TTGAGCGTCG  
 4741 ATTTCCTGTA TGCTCGTCAG GGGGGCGGAG CCTATGGAA AACGCCAGCA ACAGGGCCTT  
 4801 TTACGGTTC CTGGGCTTT GCTGGCTTT TGCTCACATG TTCTTCTG CGTTATCCCC  
 4861 TGATTCTGTG GATAACCGTA TTACCCCTT TGAGTGAAGCT GATAACCGCTC GCGCAGCCG  
 4921 AACGACCGAG CGCAGCGAGT CAGTGACCGA GGAAGCGGAA G

## Figure 7

PMSVLSB4: 6309 bp;  
 Composition 1522 A; 1620 C; 1590 G; 1577 T; 0 OTHER  
 Percentage: 24% A; 26% C; 25% G; 25% T; 0% OTHER

Molecular Weight (kDa): ssDNA: 1947.08 dsDNA: 3889.6

## ORIGIN

1	ACCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAAGCTGGC
61	ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC
121	TCACTCATTA GGCACCCCG AGCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181	TTGTGAGCGG ATAACARATT CACACAGGA ACAGCTATGA CCAATGATTAC GCCAAGCTAT
241	TTAGCTGACA CTATAGAAC TAGATGACAC CGCGCGCGAT AATTATCCT AGTTTGCGCG
301	CTAGTCCCAGA TCTAGTAACA CGTATTAAT GTATAATTGC GGGACTCTAA TCATAAAAAC
361	CTATATTTTG TTITCTATCG CGTATTAAT GTATAATTGC AATPGCTTAA CGTAATTCAA
421	CCATCTCATTA AATAACGTCG TGCAATTACAT GTTAATTATT ACATGCTTAA
481	CAGAAATTAT ATGATAATCA TCGACAGACCG CGCAACAGGA TTCAATCTTA AGAAACCTTA
541	TTGCCAAATG TTGAAACGAT CGGGGAAATT CGCTCGAGTT AATTAAGCGG CCCGCTCAA
601	AAGGATCTTC ACCTAGATCC TTTAAATTAA AAAATGAAGT TTAGCAGCTG GTCAGTCC
661	CTCTCGGGCC ACGAAGTGCA CGCAGTTGCC GGGCGGGTCG CGCAGGGCGA ACTCCCGCCC
721	CCACGGCTGC TCGCCGATCT CGGTCACTGGC CGGCCCGGAG GCGTCCCCGA AGTTGTTGA
781	CAAGACCTCC GACCACCG CGTACAGCTC GTCCAGGGCG CGCACCCACA CCCAGGCCAG
841	GGTGTGTCC GGCACCACCT GGTCTTGGAC CGCGCTGATG AACAGGTCA CGTCGCCCCG
901	GACCACACCG GCGAAGTCGT CCTCCACGAA GTCCCGGGAG AACCCGAGCC GGTGGTCCA
961	GAACTCGACC GCTCCGGGGA CGTCGCGGC GGTGAGCACC GGAACGGCAC TGGTCAACTT
1021	GGCCATGGTG GCCCTCTCA CGTGTCTTAA TTGAAGCA TTACAGGGTT ATTGTCAT
1081	GAGCGGATAC ATATTTGAAT GTATTAGAA AAATAAAACAA ATAGGGTTTC CGCGCACATT
1141	TCCCCGAAAA GTGCCACCTG TATGCGGTGT GAAATACCGC ACAGATGCGT AAGGAGAAA
1201	TACCGCATCA GGCAGAAATTG TAAACGGGGC CGCTTAATTAA AGTCGAGCTC CTCTCCAAAT
1261	GAATGAACT TCCCTATATA GAGGAAGGGT CTGCGAAGG ATAGTGGGAT TGTGGTCAT
1321	CCCTTACGTC AGTGGAGATA TCACATCAAT CCACCTGGCTT TGAAGACCTG GTTGGAACGT
1381	CTTCTTTTC CACGTAGCTC CTGCTGGGTG GGGTCCATC TTGGGACCA CTGTCGGCAG
1441	AGGCATCTTG AACGATAGCC TTTCCTTATC GCAATGATGG CAITGGTAGG TGCCACCTTC
1501	CTTTCTACT GTCTTTTGA TGAAGTGACA GATAGCTGGG CAAATGATC CGAGGAGGT
1561	TCCCGATATT ACCCTTTGTT GAAAAGTCTC AATAGCCCTT TGGCTTCTG AGACTGTATC
1621	TTTGATATTTC TTGGAGTAGA CGAGAGAGTG TCGTGTCTCA CCATGTTGAC GAATTCATGG
1681	GCAGACCCGT CTGTAATTCA AGAGTGTGG CAACAGTAA TGAATAAAAA CTCCCGTTT
1741	ATTATAATTG ATGAATGCTG AAAGCTTACA TAAATATGTC GTGGATGGC ACGAAAAAAC
1801	ACACGCAAC AATACAGGGG GGTAGTCGGC GGGCGGCTAA GGTGGTGTCT CGGCGGGCAG
1861	AACATCGAAA AATCAAGATC TATATGAATT ACACCTTCTC CGTAGGAGGA AGCACAGGG
1921	GAGAATACCA CTCTCCCCC GGCGACATAA TGAAATGAC GCAGTTGCC TCGAAATACT
1981	CCAGCTGCC TGGAGTCATT TCCCTCATCC AATCTTCATC CGAGTTGGCG AGGATTATTG
2041	TAGGCTTAGA CTCTCTGAC ACCTTTTCT TCTTACCATCA TTGGGGTTT ACAATGAAAT
2101	CCCTCTGACA GCCAACTAAC TGTTTCCAAC AAGGACAGAA TTAAACGGA ATATCATCTA
2161	CGATGTTGTA GATTGCGTCT TCGTTGTATG AAGACCAATC AACATTATTT TGCCAGTAAT
2221	TATGAACCCC TAGGCTTCTG GCCCAAGTAG ATTTCGGGT TCTTGTGGG CCGACGATGT
2281	AGAGGCTCTG CTTCCTTGAT CTTCATCTG ATGACTGGAT ACAGATCCA TCCATTGGAG
2341	GTCAGAAATT GCATCCTCGA GGGTATAACA GGTAGGTGAA AGGACCATGT AAGCTTCCGG
2401	ACTAACCTGG AAGATGTAG GCTGGAGCCA ATCGTTGATT GACTCATTAC AAAGTAAATC
2461	AGGTGAGGAG GGTGGATGAG GATTGGTGAAT CTCTTCTGAA ATCTCAGGAA AAAGTTATT
2521	TGCAGAGTAT TCAAAATACT GCAATTGTTGTT GGACCAATCA AAGGGAGCT CTTCCTGGAT
2581	CATGGAGAGG TACTCTTCTT TGGAGGTAGC GTGIGAATAA ATGTCCTGCA TTATTCATC
2641	TTTAAAGGC TTTCCTCTT TTACCTCTGA ATCAGATTTC CCTAGGAAGG GGGACTTCC
2701	AGGAATGAAA GTACCTCTCT CAAACACAGC CAGGGTTC TTGAGAATGT AATCCCTCAC
2761	TCTGTTAACT GACTTGGCAC TCTGAATT TGCGTGAAC CCATTTATAT CAAAGAACCT
2821	TGAGTCAGAT ATCGTTATCG GCTTCTCTGG CTGAAGCAAT GCATGTAAT GCAAACTTCC
2881	ATCTTATGT GCCTCTCGGG CACTAGAAAT ATATTTGGGA ATCCAACGAA CGACGAGCTC

Figure 7 (cont'd)

2941 CCAAGATCATC TGACAGGGCGA TTTCAGGATT TTCTGGACAC TTTGGATAGG TTAGGAACGT  
 3001 GTTACCGTTC CTGTGTGAGA ACTGACGGTT GGATGAGGGAG GAGGCCATAG CCGACGACGG  
 3061 AGGTTGAGGC TGAGGGATGG CAACTGGGA GCTCCAAACT CTATAGTATA CCCGTGCGCC  
 3121 TTCTGAAATCC GCGCGCTCCAT TGTCTTATAG TGGTTGIAAA TGGGCGGGAC CGGGCCGGCC  
 3181 CAGCAGGAAA AGAACGGCGC CACTAATATT ACCGCGCCCTT CTCCCCCTGC GAGGGCCGGCC  
 3241 GGTAGGGACC GAGCGCTTTG ATTTAAAGCC TGGTTCTGCT TTGTATGATT TATCTAARGC  
 3301 AGCCCAATCT AAAGAAACCG GTCCCGGGCA CTATAAATTG CCTAACAAAGT GCGATTCAATT  
 3361 CATGGATCCT TTAAACTCGA GTCTAGAGGG CCCAACCTCGC CCTATAGTGA GTCGTATTAC  
 3421 AATTCACTGG CGCTCGTTT ACAACCTCGT GACTGGAAA ACCCTGGCGT TACCCAACCTT  
 3481 AATCGCCTTG CAGCACATCC CCCTTCGCG AGCTGGCGTA ATAGCGAAGA GGCCCGCACC  
 3541 GATCGCCCTT CCCAACAGTT GCGCAGCGTA TAGTACGGC AGTTTAAGGT TTACACCTAT  
 3601 AAARGAGAGA GCGGTTATCG TCTGTTGIG GATGTACAGA GTGATATTAT TGACACGCGG  
 3661 GGGCGACCGA TGGTGTATCCC CCTGGCCAGT GCACGTCG TGTCAAGATAA AGTCTCCCGT  
 3721 GAACCTTACCG CGGTGGTGCAT TATCGGGAT GAAAGCTGGC GCATGATGAC CACCGATATG  
 3781 CGCAGTGTGC CGGTCTCCGT TATCGGGAA GAAGTGGCTG ATCTCAGCCA CGCGGAAAT  
 3841 GACATCAGAAA AGCGCAATTAA CCTGATGTTG TGGGAAATAT AAAATGTCAGG CCTGAATGGC  
 3901 GAATGGACCGC GCGCTGTAGC GCGCGCATTAA GCGCGCGGGT GTGGTGGTTA CGCGCAGCGT  
 3961 GACCGCTACA CTGCCAGCG CCCTAGCGCC CGCTCCCTTC GCTTTCTTCC CTTCTTTCT  
 4021 CGCCACGTTC GCGGGCTTTC CCCGTCAAAGC TCTAAATCGG GGGCTCCCTT TAGGTTCCG  
 4081 ATTTAGAGCT TTAACGGCACCC TCGACCGCAA AAAACTTGTAT TTGGGTGATG GTTCACGTAG  
 4141 TGGGCCATCG CCCTGATAGA CGGTTTTGCG CCCTTGTGACG TTGGAGTCCA CGTCTTTAA  
 4201 TAGTGGACTC TTGTTCCAAA CTGGAAACAAAC ACTCAACCCCT ATCGCGGTCT ATTCTTTGA  
 4261 TTTTATAAGGG ATGTTGCGGA TTTCGGCCTA TTGGTTAAAA AATGAGGTGA TTAAACAAAA  
 4321 ATTTTAACAA AATTCAAGAG AACTCGTCAA GAAGGGCGATA GAAGGGCGATG CGCTGCGAAT  
 4381 CGGGAGGGC GATAACGTAA AGCAGAGGA AGCGGTCAAGC CCATTGCGCG CCAAGCTCTT  
 4441 CAGCAATATC ACGGGTAGCC AACGGCTATGT CTCGATAGCG GTCCGCCACA CCCAGCGGCG  
 4501 CACAGCTGAT GAATCCAGAA AAGGGCCCAT TTTCACCAT GATAATTGCGC AAGCAGGCGAT  
 4561 CGCCATGGGT CACGACGGAGA TCCTCGCCGT CGGGCATGCT CGCCTTGAGC CTGCGCAACA  
 4621 GTTCGGCTGG CGCGAGCCCC TGATGCTCTT CGTCCAGATC ATCTGATCG ACAAGACCGG  
 4681 CTTCCATCGG AGTACGTGCT CGCTCGATGC GATGTTTCG TTGGTGGCTG AATGGCGAGG  
 4741 TAGCCGGATC AAGCGTATGC AGCCGCGCA TTGCACTCAGC CATGATGGAT ACTTTCTCGG  
 4801 CAGGAGCAAG GTGAGATGAC AGGAGATCCT GCCCCGGCAC TTGGCCCAAT AGCAGCCAGT  
 4861 CCCTTCCCGC TTCACTGACA ACGTCGAGCA CAGCTGCGCA AGGAACGCC GTCGTGGCCA  
 4921 GCCACGATAG CGCGCTGCG TCGCTCTGCA TTTCATTCAAG GGCACCGGAC AGGTGGTCT  
 4981 TGACAAAAAG AACCGGGCGC CCCCTGCGCTG ACAGCCGGAA CACGGCGGCA TCAGAGCAGC  
 5041 CGATTGTCIG TTGTGCCAG TCATAGCCAG ATAGCCTCTC CACCCAAAGCG GCGGGAGAAC  
 5101 CTGCGTCAA TCCATCTGT TCAATCATGC GAAACGATCC TCATCTGTC TCTTGATCAG  
 5161 ATCTTGATCC CCTGCGCCAT CAGATCCCTG GCGGGAGAA AGCCATCCAG TTACTTTGC  
 5221 AGGGCTTCCC AACCTTACCA GAGGGCGCC CAGCTGGCA TTCCGGTTCG TTGCTGCTCC  
 5281 ATAAAACCGC CCAGTCTAGC TATGCCATG TAAGGCCACT GCAAGCTACC TGCTTTCTCT  
 5341 TTGGCCTTGC GTTTCAGATAG CCCAGTAGCT GACATTCACTC CGGGGTCAAGC  
 5401 ACCGTTCTG CCGACTGGCT TTCTACGTGA AAAGGATCTA GGTGAAGATC CTTTTTGATA  
 5461 ATCTCATGAC CAAAATCCCT TAACGTCAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG  
 5521 AAAAGATCAA AGGATCTCT TGAGATCCTT TTTTCTGCG CGTAATCTGC TGCTTGCAAA  
 5581 CAAAAAAACCC ACCGCTACCA GCGGTGGTTT GTTGCGCGA TCAAGAGCTA CCAACTCTTT  
 5641 TTCCGAAAGT AACCTGGCTTC AGCAGAGCGC AGATACCAAA TACTGCTCTT CTAGTGTAGC  
 5701 CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACGCC TACATACCTC GCTCTGCTAA  
 5761 TTCTGTTACCG AGTGGCTGCT GCCAGTGGCG ATAAGCTGIG TCTTACCGGG TTGGACTCAA  
 5821 GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAAC GGGGGGTTCG TGCAACACAGC  
 5881 CCAGCTTGGA GCGAACGACCC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA  
 5941 CGGCCACGCT TCCCGAAGGG AGAAAGGGGG ACAGGTATCC GGTAAAGCGGC AGGGTGGAA  
 6001 CAGGAGAGCG CACGAGGGAG TTTCAGGGGG GAAACGCCCTG GTATCTTAT AGTCTGTCG  
 6061 GGTTTCGCA CCTCTGACTT GAGCGCTCGAT TTGTTGATG CTGTCAGGG GGGCGGAGCC  
 6121 TATGGAAAAA CGCCAGCAAC CGGGCTTTT TACGGTTCTT GGGCTTTGC TGGCTTTTG  
 6181 CTCACATGTT TTTCCTGCG TTATCCCCCTG ATTCGTGGA TAACCGTATT ACCGCCTTTG  
 6241 AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGGG CAGCGACTCA GTGAGCGAGG  
 6301 AAGCGGAAG

Figure 8

PMSVLSB-5: 8043 bp;  
 Composition 1983 A; 1992 C; 2011 G; 2057 T; 0 OTHER  
 Percentage: 25% A; 25% C; 25% G; 26% T; 0% OTHER

Molecular Weight (kDa): ssDNA: 2483.31 dsDNA: 4958.5

ORIGIN

1	AGCGCCCAAT	ACGCAAACCG	CCTCTCCCCG	CGCGTGGGCC	GATTCAATTAA	TGCAGCTGGC
61	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAAT	GTGAGCTTAGC
121	TCACTCATTA	GGCACCCCG	GCTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTTGAA
181	TTGTGAGCGG	ATAACAATT	CACACAGGAA	ACAGCTATGA	CCATGATTAC	GCCAAGCTAT
241	TTAGGTGACA	CTATAGAATA	CTCAAGCTAT	GCATCAAGCT	IGGTACCGAG	CTCGGATCCA
301	CTAGTAACGG	CCGCCAGTGT	GCTGGAATTTC	ATGGGCAGAC	CCGTCTGTAC	TTTAAGAGTG
361	TTGGCAACCA	GTAATGAATA	AAAACCTCCG	TTTATTATA	TTTGTGAAT	GCTGAAAGCT
421	TACATTAATA	TGTCGTGCGA	TGGCACGAA	AAACACACGC	AAACAATACA	GGGGGTAGT
481	CGGGGGCGG	CTAAGGGTGG	TGCTCGGCG	GCAGAACATC	GAAAAATCAA	GATCTATATG
541	AATTACACTT	CCTCCGTAGG	AGGAAGCACA	GGGGGAGAAT	ACCACTCTC	CCCCGGCGAC
601	ATAATGTAAT	TGACGCAGT	TGCCTCGAAA	TACTCCAGCT	GCCCCTGGAGT	CATTTCCTTC
661	ATCCAATCTT	CATCCGGAT	GGCGAGGATT	ATTGTAGGCT	TAGACTTCCTT	CTGCACCTTT
721	TTCCTCTTAC	CATACTGGG	TTTACAAATG	AAATCCCTCT	GACASCAAAC	TAACTGTTTC
781	CAACAAGGAC	AAATTTAAA	CGGAATATCA	TCTACGATGT	TGTAGATTGC	GTCTTCGTTG
841	TATGAAGACC	AATCAACATT	ATTTTGCCAG	TAATTATGAA	CCCTTAGGCT	TCTGGCCCAA
901	GTAGATTTC	CGGTTCTTGT	TGGGCCGACG	ATGTAGAGGC	TCTGCTTTCT	TGATCTTCA
961	TCTGTGACT	GGATACAGAA	TCCATCCATT	GGAGGTCAGA	AATTGATCC	TCGAGGGTAT
1021	AAACAGGTAGG	TTGAAGGGAGC	ATGTAAGCTT	CGGGACTAAC	CTUGAAGATG	TTAGGCTGGA
1081	GCCAAATCGTT	GATTGACTCA	TTACAAAGTA	ATTCAGGTGA	GGAGGGTGG	TGAGGATTGG
1141	TGAACCTCTC	CTGAATCTCA	GGAAAAGCT	TATTTCAGA	GTATTCAAAA	TACTGCAATT
1201	TTGTGGACCA	ATCAAAGGGG	AGCTCTTTCT	GGATCATGGA	GAGGTACTCT	TCTTGGAGG
1261	TAGCGTGTGA	AATAATGCT	CGCAATTATT	CATCTTGTAG	AGGCTTTT	TCTTACCT
1321	CTGAATCAGA	TTTCCCTIAGG	AAGGGGACT	TCTCTAGGAT	GAAAGTACCT	CTCTCAAACA
1381	CAGCCAGGAG	TICCTTGAGA	ATGTAATCCC	TCACTCTGTT	AACGTACCTG	GCACCTGTAA
1441	TATTTGGGTG	AAACCCATT	ATATCAARGA	ACCTTGAGTC	AGATATCCTT	ATCGGCTTCT
1501	CTGGCTGAG	CAATGCAATG	AAATGCAAAC	TTCCATCTTT	ATGTGCCCT	CGGGCACATA
1561	GAATATATT	GGGAATCCAA	CGAACGACGA	GCTCCCGAGAT	CATCTGACAG	GGGATTTCAG
1621	GATTTCTGG	ACACTTTGGA	TAGGTTAGGA	ACGTGTTAGC	GTTCCTGCT	GAGAAGTGC
1681	GGTTGGATGA	GGAGGGAGGCC	ATAGCCGAGC	ACGGAGGTTG	AGGCTGAGGG	ATGGCAGACT
1741	GGGAGCTCCA	AACTCTATAG	TATACCGTG	CGCCCTTCGAA	ATCCGCCGCT	CCATTGTCTT
1801	ATAGTGGTTG	AAATGGGCC	GGACGGGGCC	GGCCCAGCG	GAAAAGAAGG	CGCGCACTAA
1861	TATTACCGCG	CCTTCCTTTTC	CTGCGAGGGC	CCGGTAGGG	CCGAGCGCTT	TGATTAAAG
1921	CCTGGTCTG	CTTTGTATGA	TTTATCTAA	GCACCCCAAT	CTAAAGAAAC	CGGTCCCGGG
1981	CACTATAAT	TGCCTAACAA	GTGCGATTCA	TTCATGGATC	CTTTAAACTC	GAGTCAGTC
2041	CCGATCTAGT	AAACATAGATG	ACACCCGCC	CGATAATTAA	TCCTAGTTG	CGCGCIAATAT
2101	TTTGTCTTC	ATCGCGTATT	AAATGTATAA	TTGCGGGACT	CTAATCATAA	AAACCCATCT
2161	CATAAATAAC	GTCATGCATT	ACATGTATA	TATTACATGC	TTAACGTAAT	TCAACAGAAA
2221	TTATATGATA	ATCATCGACA	GACGGCAAC	AGGATTCAAT	CTTAAGAAC	TTTATGCCA
2281	AATGTTGAA	CGATCGGGAA	AAATCGCTCG	AGTTAATTAA	GCGGCCGCT	AAAAAGGAT
2341	CTTCACCTAG	ATCCCTTTAA	ATTTAAATG	AAAGTTTACG	ACGTGTCAGT	CTGCTCTC
2401	GGCACCGAAG	TGCACCGAGT	TGCCGGCCGG	GTCCGGCAGG	GCGAACTCCC	GCCCCCACGG
2461	CTGCTCGCG	ATCTCGGTCA	TGGCCGGCCC	GGAGGGCGTCC	CAGGAAGTTG	TGGACACGAC
2521	CTCGGACAC	TGGCGTACA	GCTCGTCCAG	GGCCGGCACC	CACACCCAGG	CCAGGGTGT
2581	GTCCGGCACC	ACCTGGTCT	GGACGGCGCT	GATGAACAGG	GTCACGTCGT	CCCGGACAC
2641	ACCGGCGAAG	TGTCCTCCA	CGAAGTCCCG	GGAGAACCCG	AGCCGGTCGG	TCCAGAACTC
2701	GACCGCTCCG	GCGACGTGCG	GGCGGGTGAG	CACCGGAACG	GCACCTGCTCA	ACTTGGCCAT
2761	GGTGGCCCTC	CTCACGTGCT	ATTATTGAAG	CTTTATTCAG	GGTTATTGTC	TCATGAGCGG
2821	ATACATATT	GAAGTATTT	AGAAAATAA	ACAAATAGGG	GTTCGGCGA	CAITTCGG
2881	AAAAGTGCCA	CCTGTTATGCG	GTGTGAATA	CCGCACAGAT	GGCTAAGGAG	AAAATACCGC

Figure 8 (cont'd)

2941 ATCAGGCAGA ATTGTAAACG CGGCCGCTTA ATTAAGTCGA CGTCCTCTCC AATGAAATG  
 3001 AACTTCCCTA TATAGAGGAA GGGTCTTGCG AAGGATAGTG GGATTGTGCG TCATCCCTTA  
 3061 CGTCAGTGGA GATATCACAT CAATCCACTT GCTTIGAAGA CGTGGTTGGA ACGTCTTCTT  
 3121 TTTCACGTA GCTCCTCGTG GGTGGGGTC CATCTTGGG ACCACTGTCG GCAGAGGCAT  
 3181 CTTGAACGAT AGCCCTTCTT TATGCCAATG ATGCCATTG TAGGTGCCAC CTTCCCTTTC  
 3241 TACTGTCCTT TTGATGAAGT GACAGATAGC TGGGCAATGG AATCCGAGGA GTTTCGGCGA  
 3301 TATTACCCCTT TGTTGAAAAG TCTCAATAGC CTTTGGTCT TCTGAGACTG TATCTTGTAT  
 3361 ATCTTGGAG TAGACGAGG AGTGTGCG TCCACCATGT TGACGAAITC ATGGCAGAC  
 3421 CCGCTGTAC TTTAAGAGTG TTGCGAACCA GTAATGAATA AAAACTCCG TTTTATTATA  
 3481 TTGATGAAT GCTGAAAGCT TACATTAATA TGTCGTGCGA TGCCACGGAA AAACACACGC  
 3541 AAACAATACA GGGGGGTAGT CGGCCGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC  
 3601 GAAAAATCAA GATCTATATG AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGAGAAT  
 3661 ACCACCTCTC CCCCCGGCAG ATAATGTAAA TGACGCACTT TGCCCTCGAAA TACTCCAGCT  
 3721 GCCCTGGAGT CATTTCCTTC ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT  
 3781 TAGACTTCTT CTGCACCTT TTCTTETTAC CATACTTGGG GTTACRAATG AAATCCCTCT  
 3841 GACAGCCAAC TAACTGTTTC CAACAAGGAC AGAATTTAA CGGAATATCA TCTACGATGT  
 3901 TGTAGATTGC GTCTTGTG TATGAAGACC ATACCAATTT ATTTCGCGAG TAATTATGAA  
 3961 CCCCTAGGCT TCTGGCCAA GTAGATTITC CGGTCTCTGT TGCCCGGACG ATGTAGAGGC  
 4021 TCTGCTTCT TGATCTTCA TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA  
 4081 AATTCATCC TCGAGGGTAT AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC  
 4141 CTGGAAAGATG TTAGGCTGGA GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA  
 4201 GGAGGGTGGG TGAGGATTGG TGAACCTTTC CTGAATCTCA GGAAAAGCT TATTGCGAGA  
 4261 GTATTCAAA TACTGCAATT TTGTTGGACCA ATCAAAAGGG AGCTCTTCTI GGATCATGGA  
 4321 GAGGTACTCT TCTTTGGAGG TAGGGTGTGA AATAATGTCT CGCATTATTT CATCTTGTAGA  
 4381 AGGCTTTTACCT TCCCTTACCT CTGAATCAGA TTTTCCCTAGG AAGGGGGACT TCTCTAGGA  
 4441 GAAAGTACCT CTCTCAAAACA CAGCCAGAGG TTCCCTTGAGA ATGTAATCCC TCACTCTGTT  
 4501 AACTGACTTG GCACTCTGAA TATTGGGTG AAACCCATTI ATATCAAGA ACCTTGAGTC  
 4561 AGATATCCTT ATCGGCTTCT CTGGCTGAAG CAATGCAATGT AAATGCAAAAC TTCCATCTT  
 4621 ATGTGCCTCT CGGGCACATA GAATATATTT GGAATTCCAA CGAACGACGA GCTCCAGAT  
 4681 CATCTGACAG GCGATTCTCG GATTTCTCG ACATTTGGA TAGTTAGGA ACGTGTTAGC  
 4741 GTTCTGTGT GAGAACTGAC GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTG  
 4801 AGGCTGAGGG ATGGCAGACT GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCCTCGAA  
 4861 ATCCGCGCGT CCATTGCTT ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCGAGCAG  
 4921 GAAAAGAAGG CGCGCACTAA TATTACCGCG CCTCTTTTC CTGCGAGGGC CGGGGGTAGG  
 4981 GACCGAGGCC TTGATTAA AGCCTGGTTC TGCTTGTAT GATTTATCTA AAGCAGCCCA  
 5041 ATCTTAAAGAA ACCGGTCCCG GGCACATATAA ATTGCCTAAC AAGTGCAGATT CATTGATGG  
 5101 TCCCTTAAAC TCGAGTCTAG AGGGCCCAAT TCGCCCTATA GTGAGTCGTA TTACAAATTCA  
 5161 CTGGCCGTG TTTTACAACG TCGTGAATGG GAAAACCTG CGGTTACCCA ACTTAATCGC  
 5221 CTGCGAGCAC ATCCCCCTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC  
 5281 CCTTCCCAAC AGTTGCGCAG CCTATACGTA CGGCAGTTA AGTTTACAC CTATAAAAGA  
 5341 GAGAGCCGTT ATCGTCTGTT TGTGGATGTA CAGAGTGTATA TTATTGACAC GCGGGGGCGA  
 5401 CGGATGGTGA TCCCCCTGGC CAGTGCACGT CTGCTGTCAAG ATAAAGTCTC CGGTGAACCT  
 5461 TACCCGGTGG TGCATATCGG GGATGAAAGC TGGCGCATGA TGACCCCGA TATGGCCAGT  
 5521 GTGCCGGTCT CGTTTATCGG GGAAGAAGTG CTGATCTCA GCCACCGCGA AAATGACATC  
 5581 AAAAAACGCA TTAACCTGAT GTTCTGGGAA ATATAATGT CAGGCCCTGA TGGCGAATGG  
 5641 ACGGCCCTG TAGCGGGCGCA TTAAGCGCGC GGGTGTGGTG GTTACCGCGA GCGTGAACCG  
 5701 TACACTTGCC AGCGCCCTAG CGGCCGCTCC TTGCGCTTTC TPCCCCTCCT TTCTCGCCAC  
 5761 GTTCCGCCGGC TTTCCCCCTC AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTAG  
 5821 AGCTTACGG CACCTCGAC GCAAAAAACT TGATTTGGGT GATGGTTCAC GTAGTGGGCC  
 5881 ATCGCCCTGA TAGACGGTTT TTGCCCCCTT GACGTTGGAG TCCACGTTCT TTAATAGTGG  
 5941 ACTCTTGTTC CAAACTGAA CAACACTCAA CCCTATCGCG GTCTATTCTT TTGATTATA  
 6001 AGGGATGTG CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTATA  
 6061 ACAAAATCA GAAGAACCTG TCAAGAACCG GATAGAACCG GATGCCGTC GAATGGGAG  
 6121 CGGCGATACC GTAAAGCAGG AGGAAGCGGT CAGCCCAITC GCCGCCAAGC TCTTCAGCAA  
 6181 TATCACGGGT AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC CGGGCACAGT  
 6241 CGATGAATCC AGAAAAGCGG CCATTTCCTA CCATGATATT CGGCAAGCAG GCATGCCAT  
 6301 GGGTCACGAC GAGATCCTCG CGTCGGGCA TGCTCGCCCTT GAGCCCTGGCG AACAGTTGG

Figure 8 (cont'd)

6361 CTGGCGCGAG CCCCTGATGC TCTTCGTCCA GATCATCCIG ATCGACAAGA CGGGCTTCCA  
 6421 TCCGAGTACG TGCTCGTCG ATGCATGTT TCGCTTGGTG GTCGAAATGGG CAGGTAGCCG  
 6481 GATCAAGCGT ATGCAGCCCG CGCATTCGAT CAGCCATGAT GGATACTTTC TCGGCAGGAG  
 6541 CAAGGTGAGA TGACAGGAGA TCCCTGCCCCG GCACCTCGC CAATAGCAGC CAGTCCCTTC  
 6601 CGCGCTTCAGT GACAACGTCG AGCACAGCTG CGCAAGGAAC GCGCGCTTCG ECAGGCCAG  
 6661 ATAGCCGCGC TGCCTCGTCT TGCAGTTCAT TCAGGGCACC GGACAGGTGCG GTCTTGACAA  
 6721 AAAGAACCGG GCGCCCGCGC GCTGACAGCC GGAACACGGC GGCAATCAGAG CAGCCGATTG  
 6781 TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA GAACCTGCGT  
 6841 GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTTGA TCAGATCTTG  
 6901 ATOCCTTGCG CCATCAGATC CTTGGCGCG AGAAAGCCAT CCAGTTTACT TTGCAAGGGCT  
 6961 TCCCAACCTT ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTGCTTGCT GTCCATAAAA  
 7021 CCGGCCAGTC TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTT CTCCTTGCGC  
 7081 TTGCGTTTC CCTTGTCCAG ATAGCCCAGT AGCTGACATT CATCGGGGT CAGCACCGTT  
 7141 TCTGCGGACT GGCTTTCTAC GTGAAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA  
 7201 TGACCAAAAT CCCTTAACGT GAGTTTCTGT TCCACTGAGC GTCAAGACCCC GTAGAAAAGA  
 7261 TCAAAGGATC TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG CAAACAAAAA  
 7321 AACCAACGGCT ACCAGCGGTG GTTGTGTTGC CGGATCAAGA GCTACCAACT CTTTTCGGA  
 7381 AGGTAACCTGG CTTCAGCAGA CGCGCAGATAC CAAATACITG CCTTCTAGTG TAGCCSTAGT  
 7441 TAGGCCACCA CTTCAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCCTGT  
 7501 TACCACTGGC TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT  
 7561 AGTTACCGGA TAAGGCGCGAG CGGTGGGCT GAACGGGGGGG TTGCTGCACA CAGGCCAGCT  
 7621 TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA GAAGGCCCA  
 7681 CGCTTCCCGA AGGGAGAAAG CGGGACAGGT ATCCGGTAAG CGGCAGGGTC GGAACACGGAG  
 7741 AGCGCACGAG GGAGCTTCA GGGGAAACG CCTGGTATCT TTATAGTCTT GTGGGTTTC  
 7801 GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGCGG AGCTATGGA  
 7861 AAAACGCCAG CAACCGGGCC TTTTACGGT TCCCTGGCTT TTGCTGGCCT TTGCTCACA  
 7921 TGGTCTTCC TGCCTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTGAGTGAG  
 7981 CTGATACCGC TCGCCGCAGC CGAACGACCG AGCGCAGCGA GTCACTGAGC GAGGAAGCGG  
 8041 AAG

Figure 9

pMSVLSB-6: 7404 bp;

Composition 1839 A; 1794 C; 1835 G; 1936 T; 0 OTHER  
 Percentage: 25% A; 24% C; 25% G; 26% T; 0% OTHER

Molecular Weight (kDa): ssDNA: 2286.33 dsDNA: 4564.5

## ORIGIN

1 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTGGGCC GATTCATTAA TGCAGCTGGC  
 61 ACGCACGGTT TCCCGACTGG AAAGCAGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC  
 121 TCACTCATTA GGCACCCCCAG GCTTTACACT TTATGCTTCC GGCTGTATG TTGTGTGAA  
 181 TTGTGAGCGG ATAACAATTTC CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT  
 241 TTAGGTGACA CTATAGAATA CTCAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA  
 301 CTAGTAACGG CCGCCAGTGT GCTGGAATTG ATGGGAGAC CCGTCTGTAC TTTAAGAGTG  
 361 TTGGCAACCA GTAATGAAATA AAAACTCCCG TTTTATTATA TTGATGAAT GCTGAAAGCT  
 421 TACATTAATA TGCGGTGCGA TGGCAGGAAA AAACACACGC AAACAATACA GGGGGTAGT  
 481 CGGGGGCGG CTAAGGGTGG TGCTCGCCG GCAGAACATC GAAAAATCAA GATCTATATG  
 541 ATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCCGGGAC  
 601 ATAATGAAA TGACCGAGTT TGCGCTGAAA TACTCCAGCT GCGCTGGAGT CAATTCTTC  
 661 ATCCAATCTT CATCCGAGTT GGGGAGGATT ATTGAGGCT TAGACTTCTT CTGCACCTTT  
 721 TTCTCTTAC CATACTGGG GTTACATAG AAATCCCTCT GACAGCCAAAC TAACTGTTTC  
 781 CAACAGGAC AGAATTAAA CGGAATATCA TCTACGATGT TGTAGATIGC GTCTTCCTG  
 841 TATGAAGACC AATCAACATT ATTTTGCAG TAATTATGAA CCCCTAGGCT TCTGGGCAA  
 901 GTAGATTTTC CGGTCTTGT TGGGCGAGG ATGAGAGGC TCTGCTTCT TGATCTTCA  
 961 TCTGATGACT GGATACAGAA TCCATCATT GGAGGTGAGA AATTGATCC TCGAGGGTAT  
 1021 AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGACTAAC CTGGAAGATG TTAGGCTGGA  
 1081 GCCAATCGTT GATTGACTCA TTACAAGTA AATCAGGTGA GGAGGGTGG TGAGGATTGG  
 1141 TGAACCTTC CIGAACTCTCA CGAAAAAGCT TATTGAGA GTATTCAAAA TACTGCAATT  
 1201 TTGIGGACCA ATCAAAGGGG AGCTCTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG  
 1261 TAGCTGTGA ATAATGTC CGCATTATT CACTTTAGA AGGCTTTTT TCCCTTACCT  
 1321 CTGAATCAGA TTTCTTAGG AAGGGGACT TCCTAGGAAT GAAAGTACCT CTCCTAAACA  
 1381 CAGCCAGAGG TTCTTGAGA ATGTAATCCC TCACCTCTT AACTGACTTG GCACTCTGAA  
 1441 TATTGGGTG AAACCCATT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATGGCTTCT  
 1501 CTGGCTGAG CAATGCATGT AAATGCAAAC TTCCATCTT ATGTGCTCTC CGGCACATA  
 1561 GAATATATTG GGGAAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG  
 1621 GATTCTCTGG ACACTTGGG TAGGTAGGA ACGTGTAGC GTTCTGTGT GAGAAGTGC  
 1681 GGTGGATGA GGAGGAGGCC ATAGCCGAGC ACGGAGGTTG AGGCTGAGGG ATGGCAGACT  
 1741 GGGAGCTCCA AACTCTATAG TATACCGTG CGCTTCTGAA ATCCGCCGCT CCATGCTTT  
 1801 ATAGGGTTG TAAATGGGC GGACCGGGC GGGCCAGAG GAAAAGAAGG CGCGCACTAA  
 1861 TATTACCGCG CCTCTCTTTC CTGCGAGGGC CGGTAGGGG CCGAGCGCTT TGATTTAAAG  
 1921 CCTGGTTCTG CTTTGTATGA TTATCTAAA GCAGCCCAAT CTAAAGAAC CGGTCCGGG  
 1981 CACTATAAT TGCCTAACAA GTGCGATTCA TTCATGGATC CTTTAAACTC GAGTCAGTC  
 2041 CCGATCTAGT ACATAGATG ACACCGGGCG CGATAATTAA TCCTAGTTG CGCGCTATAT  
 2101 TTTGTTTCT ATCGCTTAAAT AAATGTATAA TTGGGGACT CTAATCATAA AAACCCATCT  
 2161 CATAATAAAC GTCATGCTT ACATGTTAAT TATTACATGC TAAACGTAAT TCAACAGAAA  
 2221 TTATATGATA ATCATGACA GACGGGCAAC AGGATTCAAT CTTAAGAAC TTTATTGCCA  
 2281 ATGTTTGA CGATGGGAA AATTGCTCG AGTTAATTAA GCGGCCGCTT AATTAAGTCG  
 2341 ACCTCTCTC CAAATGAAAT GAACTCTT ATATAGAGGA AGGGCTTGC GAAGGATAGT  
 2401 GGGATTGTGC GTCATCCCTT ACGTCAGTGG AGATATCACA TCAATCCACT TGCTTIGAAG  
 2461 ACCTGTTCTG AACGTTCTCT TTTTCCACGT AGCTCTCGT GGGTGGGGGT CCATCTTGG  
 2521 GACCACTGTC GGCAGAGGCA TCTTGAACGA TAGCCCTTCC TTATCGCAAT GATGGCAATT  
 2581 GTAGGTGCA CCTCTCTTCT CTACTGCTCT TTGATGAAG TGACAGATAG CTGGGCAATG  
 2641 GAATCCGAGG AGGTTTCCCG ATATTACCT TTGGTAAAA GTCTCAATAG CCCTTGGTC  
 2701 TTCTGAGACT GATCTTGA TATTCTTGGG GTAGACGAGA GAGTGTGCGT CTCCACCATG  
 2761 TTGACGAATT CATGGCAGA CCCGTCGTG CTTAAGAGT GTGGCAACC AGTAATGAAAT  
 2821 AAAACTCCC GTTTATTAT ATTGATGAA TGCTGAAAGC TTACATTAAAT ATGTCGTGCG

Figure 9 (cont'd)

2881 ATGGCACGAA AAAACACACG CAAACAAATAC AGGGGGTAG TCGGCAGGGCG GCTAAGGGTG  
 2941 GTGCTCGCGC GCCAGAACAT CGAAAAATCA AGATCTATAT GAATTACACT TCCTCCGTAG  
 3001 GAGGAAGCAC AGGGGGAGAA TACCACTTCT CCCCGCGCA CATAATGTAA ATGACGCAGT  
 3061 TTGCTCGAA ATACTCCAGC TGCCCTGGAG TCATTTCTT CATCCAATCT TCATCCGAGT  
 3121 TGGCGAGGAT TATTGTAGGC TTAGACTTCT TCTGCACCTT TTCTCTTAA CCATACTTGG  
 3181 GTTTTACAAT GAAATCCCTC TGACAGCCAA CTAACTGTAA CCAACAAAGGA CAGAATTAA  
 3241 ACGGAATATC ATCTACGATG TTGTAGATTG CGTCTTCGTT GTATGAAGAC CAATCRACAT  
 3301 TATTTGCAA GTAAATTAGGAC ACCCCCTAGGC TTCTGGCCCA AGTAGATTT CCGGTTCTTG  
 3361 TTGGGCCGAC GATGTAGAGG CTCTGTTTC TTGATCTTTC ATCTGATGAC TGGATACAGA  
 3421 ATCCATCCAT TGGAGGTCAAG AAATTGCATC CTCGAGGGTA TAACAGGTAG GTGGAAGGAG  
 3481 CATGTAAGCT TCGGGACTAA CCTGGAAGAT GTAGGCTGG AGCCAAATCGT TGATTGACTC  
 3541 ATTACAAAGT AAATCAGGTG AGGAGGGTGG ATGAGGGATTG GTGAACCTCTT CCTGAATCTC  
 3601 AGGAAAAGC TTATTTGCAAG AGTATTCAAA ATACTGCAAT TTGTTGGACC AATCAAAGGG  
 3661 GAGCTTTTC TGGATCATGG AGAGGTACTC TTCTTTGGAG GTAGCGTGTG AAATAATGTC  
 3721 TCGCATTATT TCATCTTGTAG AAGGCTTTT TTCTCTTAC TCTGAAATCGA ATTTCTCTAG  
 3781 GAAGGGGAC TTCTCTAGGAA TGAAAGTACC TCTCTCAAAAC ACAGCCAGAG GTTCTTGAG  
 3841 AATGTAATCC CTCACCTCTG TAACTGACTT GGCACTCTGA ATATTTGGGT GAAACCCATT  
 3901 TATATCAAAG AACCTTGAGT CAGATATCCT TATCGGCTTC TCTGGCTGAA GCAATGCAIG  
 3961 TAAATGCAA CTTCCATCTT TATGTGCTC TCGGGCACAT AGAATATATT TGGGAATCCA  
 4021 ACGAACGACG AGCTCCCAGA TCATCTGACA GGCACTTCA GGTTTCTG GACACTTTGG  
 4081 ATAGGTTAGG AACGTGTTAG CGTTCCTG TGAGAACCTGA CGGTGGATG AGGAGGAGGC  
 4141 CATAGCCGAC GACGGAGGTT GAGGCTGAGG GATGGCAGAC TGGGAGCTCC AAACCTATA  
 4201 GTATCCCGT CGCCCTTCGA AATCCGCCGC TCCATTGCT TATAGTGGTT GTAAATGGC  
 4261 CGGACGGGC CGGGCCAGCA GGAAAAGAAG GCGCGCACTA ATATTACCGC GCCTTCTTT  
 4321 CCTGCGAGGG CGGGGGTAG GGACCGAGCG CTTGATTTA AAGCCTGGTT CTGCTTGTG  
 4381 TGATTTATCT AAAGCAGCCC AATCTAAAGA AACCGGTCCTC GGGCACTATA AATTGCTAA  
 4441 CAAGTGCAT TCATTCTATGG ATCCTTTAAA CTGGAGTCTA GAGGGCCCAA TTGCCCCAT  
 4501 AGTGAGTCGT ATTACAATTIC ACTGGCCGTC GTTTTACAAC GTGCGTACTG GAAAACCCCT  
 4561 GGCCTTACCC AACTTAATCG CCTTGCGAGCA CATCCCCCTT TCGCCAGCTG CGTAATAGC  
 4621 GAAGAGGCC GCACCGATCG CCCTTCCCAA CAGTTGCGCA GCCTATACGT ACGGCAGTTT  
 4681 AAGGTTACA CCTATAAAAG AGAGAGCCGT TATCGTCTGT TTGTTGGATGT ACAGAGTGAT  
 4741 ATTATTGACA CGCCGGGGCG ACGGATGGTGG ATCCCCCTGG CCAGTGCACG TCTGCTGTCA  
 4801 GATAAAGTCT CCGGTGAACT TTACCCCTGG GTGCGATATCG GGGATGAAAG CTGGCGCATG  
 4861 ATGACCAACCG ATATGGCCAG TGTGCGGTG TCGCTTATCG GGGAGAAGT GGCTGATCTC  
 4921 AGCCACCGCG AAAATGACAT CAAAAACGCC ATTAACCTGA TGTTCTGGG AATATAATG  
 4981 TCAGGCCCTGA ATGGCGAATG GACGCCCT GTAGCGCGC ATTAAGCGCG CGGGTGTTGG  
 5041 GGTTACGCGC AGCGTGAACCG CTACACTTGC CAGGCCCTA CGCCCGCTC TTTCGCTTT  
 5101 CTTCCCTTCC TTTCCTCGCCA CGTTGCCCG CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT  
 5161 CCCTTTAGGG TTCCGATTTA GAGCTTACG GCACCTCGAC CGAAAAAAAC TTGATTGGG  
 5221 TGATGGTTCA CGTACTGGGC CATGCCCTG ATAGACGGTT TTTCGCTCTT TGACGTTGGA  
 5281 GTCCACGTTT TTAAATAGTC GACTCTTGTG CCAAACCTGGA ACAACACTCA ACCCTATCGC  
 5341 GGTCTATTCT TTGATTTAT AAGGGATGTT GCCGATTTCG GCCTATTGGT TAAAAAAATGA  
 5401 GCTGATTAA CAAAAATTTC AACAATTC AGAAGAAACTC GTCAAGAAGG CGATAGAAGG  
 5461 CGATGCGCTG CGAACATCGGGA GCGCGATAC CGTAAAGCAC GAGGAAGCGG TCAGCCATT  
 5521 CGCCGCCAAG CTCTTCAGCA ATATCACGGG TAGCCAACGC TATGCTCTGA TAGCGGTCCG  
 5581 CCACACCCAG CGGGCCACAG TCGATGAATC CAGAAAAGCG GCCATTTC ACCATGATAT  
 5641 TCGGCAAGCA GCCATCGCCA TGGGTACCGA CGAGATCTC GCGGTGGGGC ATGCTCGCT  
 5701 TGAGGCTGGC GAACAGTTCG GCTGGCGCGA GCCCCCTGATG CTCTTCGTC AGATCATCCT  
 5761 GATCCACAAG ACCGGCTTCC ATCCGAGTAC GTGCTCGTC GATGCGATGT TTGCTTGGT  
 5821 CGTCCAAATGG GCAGGTAGGC GGATCAAGCG TATGCGAGCCG CGCATTGCA TCAGCCATGA  
 5881 TGGATACCTT CTGGCAGGA CCAAGGTGAG ATGACAGGGAG ATCCCTGCCCC GGCACCTCGC  
 5941 CCAATAGCGC CGAGTCCCTT CCGCTTCAG TGACAACTGC GAGCACAGCT GCGCAAGGAA  
 6001 CGCCCGTCGT GGCCAGCCAC GATAGCGCG CTGCTCGTC TTGCGATTC GTCAGGGCAC  
 6061 CGGACAGGTC GGTCTTGACA AAAAGAACCG GGCGCCCTG CGCTGACAGC CGGAACACGG  
 6121 CGGCATCAGA GCAGCCGATT GTCTGTTGTG CCCAGTCATA GCCGAATAGC CTCTCCACCC  
 6181 AAGCGGCCGG AGAACCTGCG TGCAATCCAT TTGTTCAAT CATCGAAAC GATCCTCATC  
 6241 CTGCTCTTG ATCAGATCTT GATCCCCCTGC GCCATCAGAT CCTTGGCGGC GAGAAAGCCA

Figure 9 (cont'd)

6301 TCCAGTTTAC TTTGCAGGGC TTCCCCAACCT TACCAAGAGGG CGCCCCAGCT GGCAATTCCG  
 6361 GTTCGCTTGC TGTCCATAAA ACCGCCAGT CTAGCTATCG CCATGTAAGC CCACTGCAAG  
 6421 CTACCTGCTT TCTCTTTCGG CTTGCGTTTT CCCTGTCCA GATAGCCAG TAGCTGACAT  
 6481 TCATCCGGGG TCAGCACCGT TTCTGCGGAC TGCGTTTCTA CGTAAAAAGG ATCTAGGTGA  
 6541 AGATCCCTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTCTG TTCCACTGAG  
 6601 CGTCAGACCC CGTAGAAAAAG ATCAAAGGAT CTCTTGAGA TCCCTTTTTT CTGCGCGTAA  
 6661 TCTGCTGCTT GCAACAAAA AAAGCACCGC TACCAAGCGGT GGTTTGTITG CCGGATCAAG  
 6721 AGCTACCAAC TCTTTTCCG AAGGTAACCTG GCTTCAGCAG AGCGCAGATA CCAAATACIG  
 6781 TCCCTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCTACAT  
 6841 ACCTCGCTCT GCTAATCCCG TTACCAAGTGG CTGCTGCCAG TGGCGATAAG TCGTGTCTTA  
 6901 CCGGGTTGGA CTCAGACGA TAGTTACCGG ATAAGGCGCA GCGGTGGGC TGAAACGGGG  
 6961 GTTCGTGCAC ACAGCCCCAGC TTGGAGCGAA CGACCTACAC CGAACCTGAGA TACCTACAGC  
 7021 GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGGGACAGG TATCCGGTAA  
 7081 GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC  
 7141 TTTATAGTCC TGTGGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTG TGATGCTCGT  
 7201 CAGGGGGGGCG GAGCCTATGG AAAAACGCCA GCAACCGGGC TTTTTACGG TTCTGGGCT  
 7261 TTTGCTGGCC TTTTGCTCAC ATGTTCTTC CTGCGTTATC CCCTGATTCT GTGGATAACC  
 7321 GTATTACCGC CTTGAGTGA GCTGATEACCG CTGCGCGCAG CGGAACGACC GAGCCAGCG  
 7381 AGTCAGTGAG CGAGGAAGCG GAAG

## SEQUENCE LISTING

<110> LARGE SCALE BIOLOGY CORPORATION  
<120> COMPOSITIONS AND METHODS FOR INHIBITING  
GENE EXPRESSION  
<130> 008010177PC00  
<140> To Be Assigned  
<141> 2001-04-04  
<150> 09/545,574  
<151> 2000-04-07  
<160> 14  
<170> FastSEQ for Windows Version 3.0  
<210> 1  
<211> 27  
<212> DNA  
<213> Cauliflower mosaic virus  
<400> 1  
tttgaattcg tcaacatggg ggagcac 27  
<210> 2  
<211> 31  
<212> DNA  
<213> Cauliflower mosaic virus  
<400> 2  
tttgtcgacg tcctctccaa atgaaaatgaa c 31  
<210> 3  
<211> 46  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> zeocin resistance gene  
<400> 3  
cccggtcgact taattaagcg gcccgcgttta caatttcgcc tgatgc 46  
<210> 4  
<211> 47  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> zeocin resistance gene  
<400> 4  
ccccctcgagt taattaagcg gcccgcctcaa aaaggatctt cacctag 47

```

<210> 5
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> nopaline synthase gene (nos) terminator sequence

<400> 5
tttctcgagc gaatttcccc gatcgttcaa ac 32

<210> 6
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> nopaline synthase (nos) terminator sequence

<400> 6
tttactagtc ccgatctagt aacatagatg ac 32

<210> 7
<211> 29
<212> DNA
<213> maize

<400> 7
tttttaatta aggtccgcct gaattctcg 29

<210> 8
<211> 30
<212> DNA
<213> maize

<400> 8
tttttaatta acggcaaggc tcacagttt 30

<210> 9
<211> 4881
<212> DNA
<213> Viral

<400> 9
agcgcccaat acgcaaaccg cctctccccg cgcggtggcc gattcattaa tgcagctggc 60
acgacaggtt tcccgactgg aaagcggca gtgagcgaa cgcaattaaat gtgagtttagc 120
tcactcatta ggcaccccgag gctttacact ttatgtttcc ggctcgatg ttgtgtggaa 180
tttgtgacgg atacaattt cacacagaa acagctatga ccatgattac gccaagctat 240
tttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca 300
ctagtaacgg cccgcgtgt gctggattc atggcagac ccgtctgtac tttaagagtg 360
ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct 420
tacattaata tgctgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt 480
cggcggccgg ctaagggtgg tgctcgccgg gcagaacatc gaaaaatcaa gatctatatg 540
aattacactt cctccgttagg aggaagcaca gggggagaat accacttctc ccccccgcac 600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gcccctggagt catttccttc 660
atccaatctt catcccgagtt ggccgaggatt attgttaggct tagacttctt ctgcaccttt 720

```

tttttcttac catacttggg gtttacaatg aaatccctt gacagccaac taactgttc	780
caacaaggac agaatttaaa cggaatatca tctacgatgt tgttagattgc gtctcggt	840
tatgaagacc aatacacatt attttgcac taattatgaa ccccttaggt tctggccaa	900
gtagatttc cggttcttgc tggggccgacg atgttagggc tctgtttct tgatcttca	960
tctgtatgtact ggatacagaa tccatccatt ggaggtcaga aattgcatt tcgagggtat	1020
aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggtgg	1080
gccaatcggtt gattgactca ttacaaagta aatcaggta ggagggtggg tgaggattgg	1140
tgaactcttc ctgaatctca ggaaaaaagct tatttgcaga gtattcaaaa tactgcaatt	1200
tttgtggacca atcaaagggg agctcttctt ggatcatggg gaggtactct tctttggagg	1260
tagcgtgtga aataatgtct cgcattattt catctttaga aggctttttt tccttacct	1320
ctgaatcaga ttttcttagg aagggggact tccttaggaat gaaagtaccc ctctcaaaca	1380
cagccagagg ttcccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa	1440
tatttgggtg aaacccattt atatcaaaga accttgagtc agatatccctt atcggcttct	1500
ctggctgaag caatgcattt aatgcacaa ttccatctt atgtgcctct cgggcacata	1560
gaatatattt gggaaatccaa cgaacgcacg gctccagat catctgacag gcgatttcag	1620
gattttctgg acactttggg taggttagga acgtgttagc gttctgtgt gagaactgac	1680
ggttggatga ggaggaggcc atagccgacg acggagggtt aggctgaggg atggcagact	1740
gggagctcca aactctatag tataccccgtg cgccctcgaa atccgcgcct ccattgtctt	1800
atagtggttt taaaatggcc ggaccggggc ggcccagcag gaaaagaagg cgccactaa	1860
tattaccgcg ccttcttttc ctgcgaggggc ccggtagggg ccgagcgctt tgattaaag	1920
cctggttctg ctttgcggcc gctcgagcat gcatcttagag ggcccaattt gccctatagt	1980
gagtcgtatt acaattcaact ggccgtcggtt ttacaacgtc gtgactgggaa aaccctggc	2040
gttacccaac ttaatcgct tcgcagcacat cccctttcg ccagctggcg taatagcgaa	2100
gaggccccca ccgatcgccc ttcccaacag ttgcgcagcc tatacgtacg gcagtttaag	2160
gtttacacct ataaaagaga gagccgttat cgtctgtttt tggatgtaca gagtgatatt	2220
attgacacgcg cgggcgacg gatggtgatc cccctggccca gtgcacgtct gctgtcgat	2280
aaagtctccc gtgaacttta cccgggtggg catatcgggg atgaaagctg gcgcatgtat	2340
accacccgata tggccagtgt gccggctcc gttatcgggg aagaagtggc tgatctcagc	2400
cacccgaaaa atgacatcaa aaacgcattt aacctgtatgt tctggggaaat ataaatgtca	2460
ggcctgaatg gcaatggac gcccctgtt gccgcgcatt aagcgcgcgg gtgtgggtt	2520
tacgcgcacg gtgaccgcta cacttgcacg ccgccttagcg cccgccttctt tgccttctt	2580
cccttcctt ctcgcacgt tcgcgggtt tccccgtcaa gctctaaatc gggggctccc	2640
tttaggggtt cgttttagag ctttacggca cctcgaccgc aaaaaacttg atttgggtga	2700
tggttcacgt agtggggccat cgcctgtata gacggtttt cgcctttgtt cgttggagtc	2760
cacgttctt aatagtggac tcttgcgttca aactggaaaca acactcaacc ctatcgccgt	2820
ctattctttt gatttataag ggtatgttgc gatttgcgc tattgtttaa aaaatgagct	2880
gatttaacaa aaatttttaac aaaattcaga agaactcgac aagaaggcga tagaaggcga	2940
tgcgtcgca atcgggagcg gcgataccgt aaagcacgag gaagcggtca gcccattcgc	3000
cgccaaatctt ttcagcaata tcacgggttgc ccaacgttat gtcctgtatag cggtccgc	3060
cacccagccg gccacagtc atgaatccag aaaaaggccc attttccacc atgatattcg	3120
gcaaggcaggc atcgccttgc gtcacgcacgac gatcctcgcc gtcgggcattt ctcgccttgc	3180
gccttggcgaa cagttcggtt ggcgcgcattt cctgtatgttgc ttcgttccagg tcacatctgtat	3240
cgacaagacc ggcttccatc cgagttacgtt ctcgcctcgat gcgatgtttt gttgtgtgtt	3300
cgaaatggca ggttgcggca tcaagcgat gcaaggccgcg cattgcataca gccatgatgg	3360
atactttctc ggcaggaggca aggtgagatg acaggagatc ctgcggccgc acttcgc	3420
atagcagcca gtccttcccc gcttcgtatca caacgtcgat cacagctgcg caaggaacgc	3480
ccgtcggttgc cagccacgtt acggcgcttgc cctcgatgttgc cagttcattt agggcaccgg	3540
acaggtcggtt ctggacaaaa agaaccgggc gcccctgcgc tgacagccgg aacacggcg	3600
catcagagca ggcattgtt tggatgttgc gtcacgcacgac gatcctcgcc tccaccaag	3660
cggccggaga acctgcgttgc aatccatctt gttcaatcat gcgaaacgtt cctcatctt	3720
tctcttgcgttgc agatcttgcgtt cccctgcgc acgatcttgc tggcgccgat aagccatcc	3780
agtttactttt gcaaggcttc ccaacccatc cagaggccgc cccagctggc aattccgggtt	3840
cgcttgcgttgc ccataaaaacc gcccagtcata gtcacgcacgac tgtaagccca ctgcacatca	3900
cctgtttctt ctttgcgttgc gctttttttt tggatgttgc gtcacatccatc tgcacatcc	3960
tccgggggtca gcaaccgttgc tgcggactgg ctttgcgttgc gaaaaggatc taggttgcgaa	4020
tcccttttgc taatctcatc accaaaatcc cttaaacgtca gttttcgatcc cactgacgtt	4080
cagaccccgat agaaaaggatc aaaggatctt cttgagatcc ttttttgc gtcgtatct	4140

gctgcttgca aacaaaaaaaaa ccaccgctac cagcggtgg tttttgccg gatcaagagc	4200
taccaactct ttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc	4260
ttcttagtcta gcccgttta ggccaccact tcaagaactc tgtagcaccc cctacatacc	4320
tcgcctctgt aatcctgtta ccagtggctg ctgcgcgtgg cgataagtctg tgtcttaccg	4380
ggttggactc aagacgatag ttaccggata aggccgcgc gtcgggctga acgggggtt	4440
cgtgcacaca gcccagctt gaggcaacga cctacaccga actgagatac ctacagcgtg	4500
agctatgaga aaggccacg cttcccgaaag ggaaaaggc ggacaggtat cccgttaagcg	4560
gcagggtcgg aacaggagag cgacgcggg agcttccagg gggaaacgcg tggtatctt	4620
atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtag	4680
ggggggcggag cctatggaaa aacgcccagca acgcggcett ttacgggtc ctgggtttt	4740
gctggcctt tgctcacatg ttcttcctg cgttatcccc tgattctgtg gataaccgta	4800
ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgacgcgagt	4860
cagtgagcga ggaagcggaa g	4881

&lt;210&gt; 10

&lt;211&gt; 3413

&lt;212&gt; DNA

&lt;213&gt; viral

&lt;400&gt; 10

agcgcacaaat acgcaaaccg cctctcccg cgcgttggcc gattcattaa tgcagctggc	60
acgacagggtt tcccgactgg aaagcggca gtgagcgc当地 cgcaattaat gtgagttac	120
tcactcatta ggcaccccgag gcttacact ttatgtttcc ggctcgatg ttgtgtggaa	180
ttgtgagcgg ataaacaattt cacacagaa acagctatga ccatgattac gccaagctat	240
tttaggtgaca ctatagaata ctcaagctat gcatcaagct tggggcccggt agggaccgag	300
cgctttgatt taaagcctgg ttctgtttt tatgattttt ctaaagcagc ccaatctaaa	360
gaaaccggtc cccggcacta taaattgcct aacaagtgcg attcattcat ggatcctta	420
aactcgcagtc tagagggccc gaattctgca gatatccatc acactggcgg cgcgtcgagc	480
atgcatctag agggccaaat tcgcccata gtgagtcgtt ttacaattca ctggccgtcg	540
ttttacaacg tcgtgactgg gaaaacctgg gcggttacccca acttaatgcg cttgcagcac	600
atcccccttt cgccagctgg cgtaatacgca aagaggcccgg caccgatcgc cttcccaac	660
agttgcgcag cctatacgta cggcagtttta aggttacac cttaaaaaaa gagagccgtt	720
atcgtctgtt tggatgtta cagagtata ttattgacac gcccggcga cggatggta	780
tccccctggc cagtcacgt ctgctgtcg ataaagtctc cctgtacact taccgggtgg	840
tgcataatcggtt ggtgaaagc tggcgtatgtt tgaccaccga tatggccagt gtgcgggtct	900
ccgttatcggtt ggaagaagtg gtcgtatca gcccggcga aatgacatc aaaaacgc当地	960
ttaacctgtat gttctggggaa atataaaatgt caggcctgaa tggcgaatgg acgcgcctg	1020
tagcggcgc当地 ttaagcgc当地 ggggtgtgg gttacgcgc当地 gctgtaccc tacacttgc当地	1080
agcgccttag cgcgcgtcc ttgcgttcc ttcccttcc ttctcgccac gttcggcggc	1140
tttccccgtc aagctctaaa tggggggcctt cctttagggg tccgatttag agcttacgg	1200
cacctcgacc gaaaaaaaact tgatttggg gatgttac gtagtggggcc atgcgc当地	1260
tagacggttt ttcgc当地 gacgttggag tccacgttct ttaatagtgg actctgttc	1320
caaactggaa caacactcaa ccctatcgcc gtctattttt ttgattata agggatgttg	1380
ccgatttccg cctattgggtt aaaaaatggg ctgatatttac aaaaattttt aaaaaattca	1440
gaagaactcg tcaagaaggc gatagaaggc gatgcgc当地 gaatggggag cggcgatacc	1500
gttaaaggcagc aggaaggcgg gccccatcc gcccggc当地 tcttc当地 gatcacgggt	1560
agccaaacgc当地 atgcctgtat agcgtccgc cacacccggc cggccacagt cgatgaatcc	1620
agaaaaggcg ccattttcca ccatgatatttcc gggcaaggc当地 gatcgccat gggtaacgc当地	1680
gagatcctcg cccgtggc当地 tgctcgccctt gagcctggcg aacagttccg ctggccgag	1740
ccccctgtatc tcttc当地 gatcatcctg atgcacaaga cccgttccca tccgagatc	1800
tgctcgatcg atgcgtatgtt tgc当地 gtc当地 gatcaaggctg gatcaaggctg	1860
atgcaggccgc cgcattcgat cagccatgat ggatacttcc tccgaggag caagggtgaga	1920
tgacaggaga tcctgccccg gcaacttccg caatagcgc当地 cagtc当地 cccgttctc	1980
gacaacgtcg agcacagctg cgcaaggaaac gcccgtcg gccaggccacg atagccgc当地	2040
tgcctcgatcg tgc当地 tcaaggccacc ggacaggctg gtcttgacaa aaagaaccgg	2100
gcccggccatc gtc当地 gcaacacggc ggc当地 cagccatgat tctgttgc当地	2160
ccagtcatag cc当地 gatccaccca agggccggaa gaacctcgct gcaatccatc	2220

ttgttcaatc atgcgaaacg atccctcatcc tgcgtcttga tcagatcttg atcccctgcg	2280
ccatcagatc cttggcgccg agaaaggccat ccagttact ttgcagggtc tcccaacatt	2340
accagagggc gccccagctg gcaattccgg ttcgttgcgt gtccataaaa ccccccagtc	2400
tagctatcgc catgttaagcc cactgcaagc tacctgtttt ctcttgcgc ttgcgtttc	2460
ccttgcgtccag atagcccagt agctgacatt catccgggtt cagcacccgt tctgcggact	2520
ggctttctac gtgaaaagaga tcttaggtaa gatctttt gataatctca tgacaaaaat	2580
cccttaacgt gagtttctgt tccactgagc gtcagacccc gtagaaaaaga tcaaaggatc	2640
ttcttgagat ccttttttc tgcgcgtaat ctgcgtcttg caaacaaaaa aaccaccgct	2700
accagcggtg gttgtttgc cggtatcaaga gctaccaact cttttccga aggttaactgg	2760
cttcagcaga gcgcagatac caaatactgt cttcttagtg tagccgtagt taggcacca	2820
cttcaagaac tctgttagcac cgcttacata cttcgctctg ctaatctgt taccagtggc	2880
tgctgcgtt ggcgataagt cgtgtcttac cgggttggac tcaagacgt agttaccgga	2940
taaggcgcag cggtcggtt gAACGGGGGGG ttcgtgcaca cagcccgact tggagcgaac	3000
gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcac cgctttccga	3060
agggagaaaag ggcgacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag	3120
ggagcttcca gggggaaacg cctggtatct ttatagttct gtcgggtttc gccacctctg	3180
actttagcgt cgattttgtt gatgctcgac agggggggcgg agcctatgaa aaaacccag	3240
caacgcggcc ttttacggt tccctggctt ttgctggctt tttgtcaca tgttcttcc	3300
tgcgttaccc cctgattctg tggataaccg tattaccggc tttgagtgag ctgataccgc	3360
tcgcccgcgc cgaacgaccc agcgcagcga gtcagtgagc gaggaagcgg aag	3413

<210> 11  
<211> 4961  
<212> DNA  
<213> Viral

<400> 11	
agcgcctaatt acgcaaaaccc cctctccccg cgcgttggcc gattcattaa tgcagctggc	60
acgacagggtt tcccgactgg aaagcggca gtgagcgcac cgcaattaaat gtgagttac	120
tcactcatta ggcaccccgag gctttacact ttatgtttcc ggctcgatgt ttgtgtggaa	180
tttgtgagcgg ataaacaattt cacacagaa acagctatga ccatgattac gccaagctat	240
tttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaaccgag ctcggatcca	300
cttagtaacgg cccggcgtgt gctggaaattc atggcagac ccgtctgtac tttaaagatgt	360
tttgcacca gtaatgaata aaaactccc ttttattata tttgtatgaat gctgaaagct	420
tacattaata tgcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggttagt	480
cgccggccgg ctaagggtgg tgctcggcg gcagaacate gaaaaatcaa gatctatatg	540
aattacactt cctccgttagg aggaagcaca gggggagaat accacttctc ccccgccgac	600
ataatgtaaa tgacgcgtt tgccctgaaa tactccagct gcccggagt cattttcttc	660
atccaatctt catccgagtt ggcgaggatt attgttagct tagacttctt ctgcaccttt	720
tttcttcttac cataacttggg gtttacaatg aaatccctt gacagccaa taactgtttc	780
caacaaggac agaattttaaa cgaaatataca tctacgtatgt ttagattgc gtcttcgttg	840
tatgaagacc aatcaacattt attttgcacg taattatgaa ccccttaggt tctggcccaa	900
gtagattttc cgggtttgtt tggggccgacg atgttagaggc tctgtttct tgatcttca	960
tctgtatgtt ggtacacaa tccatccatt ggaggtcaga aattgcatttc tcgagggtat	1020
aacaggtagg ttgaaggagc atgtaaatctt cgggactaac ctggaagatg ttaggtggaa	1080
gccaatcggtt gattgactca ttacaaatgtt aatcagggtga ggagggtgaa tgaggattgg	1140
tgaactttc ctgaatctca gaaaaaaatctt tatttgcaga gtattcaaaa tactgcattt	1200
tttgtggacca atcaaaggggg agctttttt ggatcatgaa gaggtactt tctttggagg	1260
tagcgtgtga aataatgtt cgcattatattt catctttaga aggctttttt tcctttaccc	1320
ctgaatccaga ttttcttagg aaggggactt tccttaggaat gaaagtaccc ctctcaaaca	1380
cagccagagg ttccttggaa atgtaaatccc tcactctgtt aactgacttgc gactctgaa	1440
tatttgggtt aaacccttattt atatcaaaga accttgcgtt agatatccctt atcggttct	1500
ctggctgttca agaatgttccaa ttccatctt atgtgcctt cgggcacata	1560
gaatatattt gggaaatccaa cgaacgcac gctcccgat catctgacag gcgatttcag	1620
gattttctgg acactttggaa taggttagga acgtgttgc gttccgtgtt gagaactgac	1680
ggttggatgtt gggaggaggcc atagccgacg acggagggtt aggctgaggg atggcagact	1740
gggagctcca aactctatag tataccctgt cgccttcgaa atccggccgtt ccattgttt	1800

atagtggttt	taaatgggccc	ggaccgggccc	ggcccagcag	aaaaagaagg	cgcgcactaa	1860
tattaccgcg	ccttcttttc	ctgcgagggc	ccggtaggga	ccgagcgctt	tgatttaaag	1920
cctggttctg	ctttgtatga	tttatctaaa	gcagcccaat	ctaaagaaaac	cggccccggg	1980
caactataaaat	tgccctaacaa	gtgcgattca	ttcatggatc	ctttaaactc	gagtctagag	2040
ggcccaattc	gccctatagt	gagtcgtatt	acaattcaact	ggccgtcggt	ttacaacgtc	2100
gtgactggga	aaaccctggc	gttacccaac	ttaatcgctt	tcagcacat	cccccttcgt	2160
ccagctggcg	taatagcgaa	gaggccccca	ccgatcgccc	ttcccaacacg	ttgegcagcc	2220
tatacgtacg	gcagttttaag	gtttacacct	ataaaaagaga	gagccggttat	cgtctgtttt	2280
tggatgtaca	gagtgtatatt	attgacacgc	cggggcgacg	gtatgtatcc	cccccgtggca	2340
gtgcacgtct	gtgtcgat	aaagtctccc	gtgaacttta	cccggtggtg	catatcgggg	2400
atgaaagctg	gccccatgtat	accacccgata	tggccagtgt	gccggctctcc	gttatacgggg	2460
aagaagttggc	tgatctcagc	caccgcgaaa	atgacatcaa	aaacgcattt	aacctgtatgt	2520
tctggggaat	ataaatgtca	ggcctgaatg	gcgaatggac	gcgcctgtt	gcggcgcatt	2580
aagcgccgg	gtgtgggtgt	tacgcgcagc	gtgaccgcata	cacttgcacag	cgccttagcg	2640
ccgcgtcctt	tcgttttttt	cccttectt	ctcgcacgt	tcgcccgtt	ccccgtcaaa	2700
gctctaaatc	gggggctccc	tttaggggtt	cgatttagag	cttacggca	cctcgaccgc	2760
aaaaaaacttgc	atttgggtga	tggttacgt	agtggccat	ccccctgtata	gacggggtttt	2820
cgccttttgta	cgttggagtc	cacgttctt	aatagtggac	tcttggttcca	aactggaaaca	2880
acactcaacc	ctatcgccgt	ctattttttt	gattataag	ggatgttgcc	gatttcggcc	2940
tattggtaa	aaaatgagct	gattdaaca	aaattttaaac	aaaattcaga	agaaactcgtc	3000
aagaaggcga	tagaaggcga	tgcgctgcga	atcgggagcg	gcgataaccgt	aaagcacag	3060
gaagecggtca	gcccattcgc	cgccaagcgc	ttcagcaata	tcacgggttag	ccaaacgctat	3120
gtcctgatag	cggccgcaca	caccgcgccc	gccacagtcg	atgaatccag	aaaagcggcc	3180
attttccacc	atgatattcg	gcaaggccggc	atcggcatgg	gtcacgcacga	gatcctcgcc	3240
gtcgggcatg	ctgccttgta	gcctggcgaa	cagttcggt	ggcgcgagcc	cctgtatgtc	3300
ttcgtccaga	tcatcctgtat	cgacaagacc	ggcttccatc	cgagtcacgt	ctcgctcgat	3360
gcgatgttgc	gcttggtgtt	cgaatggca	ggtagccgga	tcaagcgtat	gcagccgccc	3420
cattgcatca	gcccattatgg	atactttctc	ggcaggagca	aggtgagatg	acaggagatc	3480
ctgccccggc	acttcgccta	atagcagcca	gtccccctccc	gttcaagtgt	caacgtcgag	3540
cacagctgcg	caaggaacgc	ccgtcggtgc	cagccacgat	agccgcgtgc	cctcgcttt	3600
cagtccatc	agggcaccgg	acaggtcggt	tttgacaaaa	agaaccgggc	ccccctgcgc	3660
tgacagccgg	aacacgcgg	catcagagca	gccgattgtc	tgttgcgcc	agtcatagcc	3720
gaatagccgc	tccacccaag	cgcccgaggaa	acccgtgtc	aatccatctt	gttcaatcat	3780
gcgaaacgt	cctcattctgt	tcttgcgtat	agatcttgc	ccccctgcgc	atcagatctt	3840
tggccgcgag	aaagccatcc	agtttactt	gcaggcttc	ccaaacettac	cagaggcgc	3900
cccaagctggc	aattcccggt	cggttgcgt	ccataaaaacc	gcccaactca	gtatcgcca	3960
tgttaaggcca	ctgcaagcta	cctgtttttt	ctttgcgtt	gcgtttttcc	ttgtccagat	4020
agccccatgt	ctgacattca	tccgggggtca	gcaccgtttc	tcggactgg	cttctacgt	4080
aaaaagatc	tagtgaaga	tccttttttga	taatctcatg	accaaaaatcc	tttaacgtga	4140
gttttgcgtt	cactgagcgt	cagacccgt	agaaaagatc	aaaggatctt	tttgagatcc	4200
ttttttctg	cgcgtatct	gtgtttgc	aaaaaaaaaa	ccaccgcctac	cagcggtgg	4260
tttgttgcgc	gatcaagagc	taccaactt	ttttccgaag	gttaactggc	tcagcagagc	4320
gcagatacca	aataactgtcc	ttcttagtgc	gccgtatgtt	ggccaccact	tcaagaactc	4380
tgttagcaccc	cctacatacc	tcgtctgtt	aatctgtt	ccagttggctg	ctggcagttt	4440
cgataagtcg	tgtttaaccg	ggttggactc	aagacgtat	ttaccggata	aggcgcagcg	4500
gtcgggcgtga	acgggggggtt	cgtgcacaca	gcccgatctt	gagcgcacga	cctacaccga	4560
actgagatac	ctacagcggt	agctatgaga	aagcgcacag	tttcccgaag	ggagaaaaggc	4620
ggacaggtat	ccgttaagcg	gcagggtcg	aacaggagag	cgcacgcggg	agcttccagg	4680
gggaaacgcgc	ttgttatctt	atagtcgtt	cggttttcgc	cacctctgac	ttgagcgtcg	4740
atttttgtga	tgctcgat	gggggcggag	cctatggaaa	aacgcgcacga	acgcggccctt	4800
tttacgttgc	ctggcccttt	gttgcgtt	tgctcatacg	tttcttctgt	cgttatacccc	4860
tgattctgtg	gataaccgtt	ttaccgcctt	tgagtgcgt	gataccgcgc	gccgcagccg	4920
aacgaccgg	cgcagcgagt	cagtgcgtca	ggaagcgaa	g		4961

&lt;210&gt; 12

&lt;211&gt; 6309

&lt;212&gt; DNA

&lt;213&gt; Viral

&lt;400&gt; 12

agcgcccaat acgaaaacccg cctctccccg	cgcgtggcc gattcattaa	tgcagctggc	60
acgacaggtt tcccactgg aaagcggca	gtgagcgc当地	cgcaattaat	120
tcactcatta ggcaccccaag	gcttacact ttatgctcc	ggctcgatg	180
tttgagcg	ataacaattt cacacaggaa	acagctatga	240
tttagtgaca	ctatagaata ctcaagctat	gcatcaagct	300
ctagtccca	tctagtaaca tagatgacac	cgcgccgat	360
ctatattttg	ttttctatcg cgtattaaat	gtataattgc	420
ccatctata	aataacgtca tgcattacat	ggactctaa	480
cagaaattat	atgataatca tcgacagacc	ttcaatctta	540
ttgccaaatg	tttgaacgat cggggaaatt	cgctcgagtt	600
aaggatctc	acctagatcc tttaaatta	aattaagcgg	660
ctcctcgcc	acgaagtgc	cgcagttgcc	720
ccacggctgc	tcgccgatct	cggtcatggc	780
cacgaccc	gaccactcg	cgtagcgtc	840
ggtgttgc	ggcaccacct	gttctggac	900
gaccacaccg	gcgaagtgc	cgcgctgatg	960
gaactcgacc	cctccacgaa	gtccccggag	1020
ggccatggtg	gctccggcga	cgtcgcgc	1080
cccttacgtc	cggtgatct	gggtccggag	1140
gagcgatata	atattgaat	gtatggat	1200
tccccaaaa	gtgcacactg	tatgcggtgt	1260
taccgcata	ggcgaatttgc	taaacgcgca	1320
gaaatgaact	tccttatata	gaggaagggt	1380
cccttacgtc	cttgcgtat	ttgaagcatt	1440
tttcttttc	tttgcgtat	tatcagggtt	1500
aggcatcttgc	tttgcgtat	tttgcgtat	1560
cttttctact	tttgcgtat	tttgcgtat	1620
tcccgcattt	tttgcgtat	tttgcgtat	1680
tttgcattt	tttgcgtat	tttgcgtat	1740
gcagaccgt	tttgcgtat	tttgcgtat	1800
attatatttgc	tttgcgtat	tttgcgtat	1860
acacgcacaa	tttgcgtat	tttgcgtat	1920
aacatcgaaa	tttgcgtat	tttgcgtat	1980
gagaataccca	tttgcgtat	tttgcgtat	2040
ccagctgccc	tttgcgtat	tttgcgtat	2100
taggctttaga	tttgcgtat	tttgcgtat	2160
ccctctgaca	tttgcgtat	tttgcgtat	2220
cgatgttgc	tttgcgtat	tttgcgtat	2280
tatgaaccc	tttgcgtat	tttgcgtat	2340
agaggctctg	tttgcgtat	tttgcgtat	2400
gtcagaaatt	tttgcgtat	tttgcgtat	2460
actaacctgg	tttgcgtat	tttgcgtat	2520
tttgcgtat	tttgcgtat	tttgcgtat	2580
tttgcgtat	tttgcgtat	tttgcgtat	2640
tttgcgtat	tttgcgtat	tttgcgtat	2700
tttgcgtat	tttgcgtat	tttgcgtat	2760
tttgcgtat	tttgcgtat	tttgcgtat	2820
tttgcgtat	tttgcgtat	tttgcgtat	2880
tttgcgtat	tttgcgtat	tttgcgtat	2940
tttgcgtat	tttgcgtat	tttgcgtat	3000
tttgcgtat	tttgcgtat	tttgcgtat	3060
tttgcgtat	tttgcgtat	tttgcgtat	3120
tttgcgtat	tttgcgtat	tttgcgtat	3180
tttgcgtat	tttgcgtat	tttgcgtat	3240

ggtagggacc gagcgctttg atttaaagcc tggttctgct ttgtatgatt tatctaaagc	3300
agcccaatct aaagaaaaccc gtcggggca ctataaattt cctaacaagt gcgattcatt	3360
catggatccc ttaaactcgaa gtctagaggg cccaaatccgc cctatagtga gtcgtattac	3420
aattcactgg ccgtcggttt acaacgtcggt gactggaaa accctggcgta tacccaactt	3480
aatcgccctt cagcacatcc cccttgcgc agctggcgta atagcgaaga ggccccgacc	3540
gatcgccctt cccaaacagtt ggcgagccctt tacgtacggc agtttaagggt ttacacctat	3600
aaaagagaga gcccgttatcg tctgtttgtg gatgtacaga gtgatattat tgacacgccc	3660
ggcgacgga tggtgatccc cctggccagt gcacgtctgc tgcagataa agtctccgt	3720
gaactttacc cggtggcgca tatecgggat gaaagctggc gcgtatgac caccgatatg	3780
gccagtgtgc cggtctccgt tatecgggaa gaagtggctg atctcagccca cccgaaaaat	3840
gacatcaaaa acgcattaa cctgtatgtc tggggatataa aatgtcagg cctgaatggc	3900
gaatggacgc gcccgttgc ggcgcattaa gcgcgcgggt gtgggtggta cgccgcacgt	3960
gaccgctaca cttggccagcg ccctagcgcc cgctccccc gcgttcttcc cttccttct	4020
cgcacatggc gccgcttcc cccgtcaagc tctaatacg gggctccctt tagggttccg	4080
attttagagct ttacggcacc tcgaccgcaa aaaacttgc ttgggtgtat gttcacgtag	4140
tgggcattcg ccctgtataga cgggttttcg cccttgcacg ttggagtcgc cgttctttaa	4200
tagtggactc ttgttccaaa ctggaaacaac actcaaccct atcgcggctc attctttga	4260
tttataaggg atgtgcccga ttccggctt ttgggtaaaa aatgagctga ttaacaaaaa	4320
attttaacaa aattcagaag aactcgtaa gaaggcgata gaaggcgatg cgctgcgaat	4380
cgggagccgc gataccgtaa agcacgagga agcggtcagc ccattcgccg ccaagctctt	4440
cagcaatatac acgggttagcc aacgctatgt cctgtatgcg gtccgcacaca cccagccggc	4500
cacagtcatgaaatccagaa aagcggccat ttccaccat gatattcgac aagcaggcat	4560
cgcacatggc cacgacgaga tcctcgccgt cgggcacatgc cgccttgcgc ctggcgaa	4620
gttcggctgg cgcgagccccc tgatgtctt cgtccagatc atcctgtatgc acaagaccgg	4680
cgttccatccg agtacgtgct cgctcgatgc gatgttgcg ttgggtggc aatgggcagg	4740
ttagccggatc aagcgtatgc agccgcgcga ttgcacatgc catgtatggat actttctcg	4800
caggagcaag gtgagatgac aggagatctt gccccggcacttccgcataat agcagccagt	4860
ccctcccccgttccatgtac acgtcgacca cagctcgca aggaacgcgc gtcgtggca	4920
gccacgatag ccgcgtcgcc tcgttgcgc ttgcattcag ggcacccggc aggtcgctc	4980
tgacaaaaaa aaccggccgc ccctcgccgt acagccggaa cacggcgccca tcagacgc	5040
cgattgtctg ttgtgcccag tcatacgccatc atagcccttc caccacatgc gccggagaac	5100
ctgcgtgcacccatctgt tcaatcatgc gaaacatcc tcacatcgatc tcttgcata	5160
atcttgcattcc cctgcgcattc cagatccctg gcccggagaa agccatccatc tttactttgc	5220
agggtttccc aacccatccca gaggggccgc cagctggcaat ttccgggttcg cttgcgttcc	5280
ataaaaacccgc ccagcttagc tatacgccatc taagccact gcaagctacc tgctttctct	5340
ttgcgtgcacccatctgt gtcggatccatc cccagatgc gacattcata cgggggtcagc	5400
accgtttctg cggactggctt ttctacgtca aaaggatcta ggtgaagatc ctttttgcata	5460
atctcatgac caaaatccct taacgtgatc ttgcattcgc ctgagcgtca gaccccgtag	5520
aaaagatcaa aggatcttgc tgagatccctt tttttctgcg cgtaatctgc tgcttgc	5580
caaaaaaaaaacc accgttacca gcggtggtt gtttgcggca tcaagatgc ccaactctt	5640
ttccgaaaggtaaactggctt acgcagacgcg agataccataa tactgtccctt ctatgttagc	5700
cgtatgttccgatccacttcc aagaactctg tagcaccgc tacataccctc gctctgttcc	5760
tcctgttacc agtggctgtc gccagtgccg ataagtcgtg tcttaccggg ttggactcaa	5820
gacgatagtt accggataag ggcgacgcgtt cgggcgttac ggggggttcg tgcacacac	5880
ccagcttggc gcgaaacgcacc tacaccgaaatc tgagatccatc acagcgttagt ctatgagaaa	5940
gcccacgcgttcccgaaaggaaaggccgc acaggatcc ggtaaacgcgc agggtcgaa	6000
caggagacgc caccggggatc ttcccgaggaa gaaacgcgttac gttatcttgc tggactcaa	6060
ggtttgcgcacccatctgtt gtcggatccatc ttttgcata cgggggttcg gggggagcc	6120
tatggaaaaaa cgcacatccca gcgcccttcc tccatcccttgc tggcttttgc tggccttttgc	6180
ctcacatgtt cttccatccca ttatcccttgc attctgttgc taaccgtatt accgccttttgc	6240
agtggatgtca taccgtatgc cgcacgcgaa cgaccgacgcg cagcgtatgc gtgagcggagg	6300
aagcggaaag	6309

<212> DNA  
 <213> Viral

<400> 13

agcgcccaat acgcaaaccg cctctccccg cgcggtggcc gattcattaa tgcagctggc	60
acgacaggtt tcccactgg aaagcgggca gtgagcgcaa cgcattaat gtgagtttagc	120
tcaactttaa ggcaccccaag gtttacact ttatgcttc ggctcgatg ttgtgtggaa	180
ttgtgagcgataacaattt cacacaggaa acagctatga ccatgattac gccaagctat	240
tttaggtgaca ctatagaata ctaaagctat gcatcaagct tggtaaccgag ctcggatcca	300
ctagtaacgg cgcggagtgt gtttgcattc atggcagac ccgtctgtac tttaaagatg	360
ttggcaacca gtaatgaata aaaactcccg ttttattata ttgtatgaat gctgaaaagct	420
tacattaata tgcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggttagt	480
cggcggccgg ctaagggtgg tgctcgccg gcagaaacatc gaaaaatcaa gatctatatg	540
aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccgccgac	600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc	660
atccaatctt catccgagtt ggcgaggatt attgttaggt tagacttctt ctgcaccttt	720
ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaaac taacttttc	780
caacaaggac agaatttaaa cggaatatca tctacatgt ttagatgtgc tgtagattgc	840
tatgaagacc aatcaacattt attttgcacg taattatgaa ccccttaggtc tctggcccaa	900
gttagatttc cggttcttgc tggggccgacg atgttagaggc tctgtttct tgatcttca	960
tctgtatgact ggatacagaa tccatccatt ggaggtcaga aattgcatttc tcgagggat	1020
aacaggtagg ttgaaggagc atgtaaagctt cgggactaaac ctggaaagatg ttaggctgga	1080
gccaatcggtt gattgactca ttacaaagta aatcaggtga ggagggttggc tgaggattgg	1140
tgaacttttc ctgaatctca ggaaaaagct tatttgcaga gtattcaaaa tactgcattt	1200
ttgtggacca atcaaagggg agcttcttgc ggatcatggg gaggtactct tctttggagg	1260
tagcgtgtga aataatgtct cgcattattt catctttaga aggctttttt tcttttacct	1320
ctgaatcaga ttttcttagg aagggggact tccttaggaat gaaagtacct ctctcaaaaca	1380
cagccagagg ttcccttgaga atgtaaatccc tcactctgtt aactgacttg gcactctgaa	1440
tatttgggtg aaaccctattt atatcaaaga accttgcgtt agatatcctt atcggcttct	1500
ctggctgaag caatgcattt aatgcattt ttccatcttt atgtgcctct cgggcacata	1560
gaatatattt gggatccaa cgaacgcacgc gctcccgat catctgcacag gcgatttcag	1620
gattttctgg acactttggaa taggttaga acgtgttagc ttctgtgtg gagaactgac	1680
ggttggatgaa ggaggaggcc atagccgacg acggagggttgg aggtcgaggat atggcagact	1740
gggagctcca aactctatag tatacccgtg cgcctcgaa atccggcgct ccattgtctt	1800
atagtgggtt gaaatggggcc ggaccggggc ggcccagcag gaaaagaagg cgcgcactaa	1860
tattaccgcg ccttcttttc ctgcgaggc cccgttagggc cccggcgtt tgattttaag	1920
cctgggtctg ctttgtatgaa ttatctaa gcagcccaat ctaagaaac cggtcccggg	1980
cactataat tgcctaacaat gtgcgattca ttcatggatc ctttaaactc ggtctagtc	2040
ccgatctgtt aacatagatg acaccgcgc cgcataattt tccttagtttgc cgcgttat	2100
tttgtttctt atgcgttataa aatgtataa ttgcgggact ctaatcataa aaaccatct	2160
cataaataac gtcatgcattt acatgttataa tattacatgc ttaacgtataa tcaacagaaa	2220
ttatgtatgata attcgcataa gaccggcaac aggattcaat ctaagaaac ttattgtcca	2280
aatgtttgaa cgcattggggaa aattcgctcg agttaattaa gcggccgcct caaaaaggat	2340
cttcacccatggatccatccatggatccatccatggatccatccatggatccatccatggatccat	2400
ggccacgcacgc tgcacgcgtt tgcacgcgtt tgcacgcgtt tgcacgcgtt tgcacgcgtt	2460
ctgcgtcgccg attcgcgttca tggccggccccc ggaggcggtcc cggaaatgtcg tggacacgc	2520
ctccggccacac tggcgatcgttca gtcgttccatccatggatccatccatggatccatccatggatccat	2580
gtccggccacc acctgggtcttccatccatggatccatccatggatccatccatggatccatccatggatccat	2640
accggcggaaatggatccatccatggatccatccatggatccatccatggatccatccatggatccatccatggatccat	2700
gacccgtccatccatggatccatccatggatccatccatggatccatccatggatccatccatggatccatccatggatccat	2760
gggtggcccttc ctacacgtgttca attattgtatgaa cattttatcgat gtttattgttgc tcatgagcgg	2820
atacatattt gaatgtatattt aaaaaaaaaaa acaaataatggg ttcccgccgc cattttcccg	2880
aaaagtggca cctgtatgcgtt gttgtatggaaataa cccgcacat gctgtatggaaatgggaaaataccgc	2940
atcaggcgaa attgtatggaaatgggaaaataccgc cggccgcgttca attaagtgcgttca ctttgcgttca aatgtatggaaatgggaaaataccgc	3000
aacttcctta tataaggaggaa ggggttttgcgttca aaggatgttgcgttca ggttttgcgttca aacttcctta	3060
cgtcgttgcgttca gatatcacat ctttgcgttca gttttgcgttca gttttgcgttca gttttgcgttca	3120
tttccacacgttca gtcctcgatccatccatggatccatccatggatccatccatggatccatccatggatccatccatggatccat	3180

cttgaacat	agcccccct	tatcgcaatg	atggcatttgc	taggtgccac	cttccttttc	3240
tactgtcctt	tttgatgaagt	gacagatagc	tgggcaatgg	aatccgagga	ggtttcccgaa	3300
tattaccctt	tgttggaaaag	tctcaatagc	ccttgggtct	tctgagactg	tatctttgat	3360
attcttggag	tagacgagag	agtgtcggtc	tccaccatgt	tgacgaattc	atgggcagac	3420
ccgtctgtac	tttaagagtg	ttggcaaccac	gtaatgaata	aaaactcccg	ttttattata	3480
tttgcgttat	gctgaaaagct	tacattaata	tgtcgtgcga	tggcacgaaa	aaacacacgc	3540
aaacaataca	ggggggtagt	ccggccccgg	ctaagggtgg	tgtcggccgg	gcagaacatc	3600
aaaaaatcaa	gatctatatg	attacactt	cctccgtagg	agaaggcaca	ggggggagaat	3660
accacttctc	ccccggcgac	ataatgtaaa	tgacgcgtt	tgcctcgaaa	tactccagct	3720
gcctggagt	catttccttc	atccaatctt	catccgagtt	ggcgaggatt	attgttaggct	3780
tagacttctt	ctgcacccctt	ttcttcttac	catacttggg	gttacaatg	aaatccctct	3840
gacagccaaac	taactgtttc	caacaaggac	agaattttaa	cggaaatatca	tctacgatgt	3900
tgtagattgc	gtctcggtt	tatgaagacc	aatcaacatt	attttgcag	taattatgaa	3960
cccccttaggt	tctggcccaa	gtagattttc	cgggttctgt	tggggccgacg	atgttagaggc	4020
tctgccttct	tgatctttca	tctgtatgact	ggatacagaa	tccatccatt	ggaggtcaga	4080
aattgcattcc	tcgagggat	aacaggtagg	ttgaaggagc	atgtaaatctt	cgggactaac	4140
ctggaaagatg	tttaggcttgc	gcataatcggt	gattgactca	ttacaaagta	aatcagggtga	4200
ggaggggttgg	tgaggattttg	tgaactcttc	ctgaatctca	ggaaaaagct	tatttgcaga	4260
gtattcaaaa	tactgcaattt	ttgtggacca	atcaaagggg	agctttttct	ggatcatgga	4320
gaggacttct	tctttggagg	tagcgtgttgc	aataatgtct	cgcattattt	catctttaga	4380
aggctttttt	tcctttacct	ctgaatcaga	ttttcttagg	aaggggggact	tccttaggaat	4440
gaaagttaccc	ctctcaaaca	cagccagagg	ttccttgaga	atgtaaatccc	tcactctgtt	4500
aactgacttg	gcactctgaa	tatttgggtt	aaaccattt	atataaaga	acctttagtgc	4560
agatatcctt	atcggttct	ctggctgaaag	caatgcattt	aatatgcac	ttccatctt	4620
atgtgcctt	cgggcacata	gaatataattt	gggaatccaa	cgaacgacga	gctcccgat	4680
catctgacag	gcgatttcag	gattttctgg	acactttgg	taggttagga	acgtgttagc	4740
gttcctgtgt	gagaactgac	ggttggatga	ggaggaggcc	atagccgacg	acggagggtt	4800
aggctgaggg	atggcagact	gggagcttca	aactctatag	tataccctgt	cgccttcgaa	4860
atccgcctt	ccattgttctt	atagtgggtt	taaatggggc	ggaccggggc	ggcccgacg	4920
gaaaagaagg	cgcgcactaa	tattaccgcg	ccttcttttc	ctgcgaggggc	ccggggtagg	4980
gaccgagcgc	tttgattttaa	agcctgggtc	tgctttgtat	gatttatcta	aagcagccca	5040
atctaaagaa	accggccccg	ggcactataa	attgcctaa	aagtgcgatt	cattcatgga	5100
tcctttaaac	tcgagtcttag	agggcccaat	tcgcctata	gtgagtcgta	ttacaattca	5160
ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	gcttacccca	acttaatcg	5220
cttgcagcac	atccccctt	cgccagctgg	cgtaatagcg	aagaggcccc	caccgatcg	5280
ccttcccaac	agttgcgcag	cctatacgt	cggcagttt	agtttacac	ctataaaga	5340
gagagccgtt	atcgctctgtt	tgtggatgt	cagagtata	ttatgtacac	gccggggcga	5400
cggtatggta	tccccctggc	cagtgcacgt	ctgctgtcag	ataaaatctc	ccgtgaactt	5460
tacccgggtt	tgcataatcg	ggatgaaagc	tggcgcata	tgaccacca	tatggccagt	5520
gtgcccgtt	ccgttatcg	ggaagaatgt	gtgtatctca	gcccacgcga	aaatgacatc	5580
aaaaacgcac	ttaacctgtat	tttctgggg	atataatgt	caggcctgaa	tggcgaatgg	5640
acgcgcctt	tagcgccgc	ttaagcgcgc	gggtgtgggt	gttacgcgc	gcgtgaccgc	5700
tacacttgc	acgcgcctt	cgcggcgtcc	tttcgttctt	ttcccttcct	ttctcgccac	5760
gttcggccgc	tttccccctgtc	aaatcttca	tcgggggctc	cttttaggtt	tccgatttag	5820
agctttacgg	cacettcgacc	gaaaaaaaaact	tgattttgggt	gatgggttac	gtagtggcc	5880
atgcgcctgt	tagacgggtt	ttcgccttt	gacgttggag	tccacgttct	ttaatagtgg	5940
actcttgc	caaactggaa	caacactcaa	ccctatcg	gtctattttt	ttgatttata	6000
agggatgtt	ccgatttcgg	cctattttgtt	aaaaaatgag	ctgatataac	aaaaattttt	6060
acaaaattca	gaagaactcg	tcaagaaggc	gatagaaggc	gatgcgtgc	gaatcgggag	6120
cggcgatacc	gtaaaagcagc	aggaaggcggt	cagcccatc	gccgccaagc	tcttcagca	6180
tatcacgggt	agccaaacgt	atgtcctgtat	agcgttccgc	cacaccacgc	cggccacagt	6240
cgatgaatcc	agaaaaagcg	ccatcttca	ccatgatatt	cgcaagcag	gcatcgccat	6300
gggtcacgac	gagatcctcg	ccgtcgccgg	tgctcgccctt	gagcctggcg	aacagttcg	6360
ctggcgccgag	ccccctgtatgc	tcttcgttca	gatcatcctg	atcgacaaga	ccggcttcca	6420
tccgagtgac	tgctcgctcg	atgcgtatgtt	tcgctttgggt	gtcgaatggg	caggttagcc	6480
gatcaaggcg	atgcagccgc	cgcatcgat	cagccatgtat	ggatactttt	tcggcaggag	6540
caaggtgaga	tgacaggaga	tcctggcccg	gcacttcg	caatagcagc	cagtcccttc	6600

ccgcttcagt gacaacgtcg agcacagctg cgcaaggaac gcccgtcgtg gccagccacg	6660
atagccgcgc tgcctcgtct tcagggcacc ggacaggtcg gtcttgacaa	6720
aaagaacccgg ggcgccttcg gctgacagcc ggaacacggc ggcattcagag cagccgattg	6780
tctgttgc ccagtcatag cggaatagcc tctccaccca agcggccggaa gaacctgcgt	6840
gcaatccatc ttgttcaatc atgcgaaacg atcctcatcc tgtcttttga tcagatcttgc	6900
atccccgtcg ccatcatc cttggccggcg agaaagccat ccagttact ttgcagggt	6960
tcccaacccctt accagaggc gcccagctg gcaattccgg ttcgcttgcgt gtccataaaa	7020
ccgcccagtc tagctatcgc catgtaaagcc cactgtcaagc tacctgtttt ctctttgcgc	7080
ttgcgttttc ccttgtccag atagcccagt agctgacatt catccgggtt cagcaccgtt	7140
tctgcggact ggcttctac gtgaaaaggaa tctaggtaa gatcctttt gataatctca	7200
tgacaaaaat ccctaaccgt gagtttctgt tccactgac gtcagaccccc gtagaaaaaaga	7260
tcaaaggatc ttcttgcgtat ctttttttc tgcgctgtat ctgctgtttt cttttttccga	7320
aaccaccgtt accagcgggtg gtttgcgttgc cggatcaaga gcttccact cttttttccga	7380
aggttaactgg cttcagcaga ggcgcagatac caaaatctgt ctttcttagt tagccgtatg	7440
taggccacca cttaagaac tctgttagcac cgcctacata cttcgctctg ctaatctgt	7500
taccagtggc tgctgcgtat ggcgataagt cgtgttttac cgggttggac tcaagacgtat	7560
agttaccggta taaggcgcag cgggtgggtt gaaacgggggg ttcgtgcaca cagcccgat	7620
tggagegaaac gacccatcacc gaaactgagat acctacagcg tgagctatga gaaagcgcac	7680
cgttcccgaa agggagaaaag ggggacaggat atccgttaag cggcagggtt ggaacaggag	7740
agcgacacgag ggagcttcca gggggaaacg cctggatct ttatgttctt gtcgggtttc	7800
gcccacccctg acttgcgtt cgtttttgt gatgctgcgtc agggggggcg gacctatggaa	7860
aaaacgcacccaaac caacgcggcc tttaacggt tccctggctt ttgctggctt tttgctcaca	7920
tgttctttcc tgctgttatec cctgttttgc tggataaccg tattaccggc tttgagtgag	7980
ctgataccgc tcgcccgcgc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg	8040
aag	8043

&lt;210&gt; 14

&lt;211&gt; 7404

&lt;212&gt; DNA

&lt;213&gt; Viral

&lt;400&gt; 14

agcgcccaat acgcaaaccg cctctccccg cgcggtggcc gattcattaa tgcagctggc	60
acgacaggtt tcccgactgg aaagcggca gtgagcgcggaa cggaaattaaat gtgagttac	120
tcactcatta ggcaccccgag gcttttactt ttatgttcc ggctcgatgt ttgtgtggaa	180
ttgtgagccg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat	240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaaccgg ctcggatcca	300
cttagtaacgg cggccaggtt gcttggaaatc atggcggac cggctctgtac tttaaagatgt	360
ttggcaacca gtaatgaata aaaactcccg ttttattata ttgtatgaat gctgaaagct	420
tacattaata tgcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggttagt	480
cggcggccgg ctaagggtgg tgctcgccgg gcagaaacatc gaaaaatcaa gatctatatg	540
aattacactt cctccgtagg aggaaggcaca gggggagaat accacttctc ccccgccgac	600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gcccctggagt catttccttc	660
atccaatctt catccgagtt ggcgaggatt attgttaggt tagacttctt ctgcaccttt	720
ttcttcttac catacttggg gtttacaatg aaatccctt gacagccaaac taacttttc	780
caacaaggac agaatttaaa cggaaatataca tctacatgt ttagatgtc gtcttcgttg	840
tatgaagacc aatcaacattt attttgcacg taattatgaa ccccttaggtt tctggccaa	900
gtatgtttc cgggttcttgc tggggccgacg atgttagggc tctgtttct ttagtcttca	960
tctgtatgtact ggatacagaa tccatccatt ggaggtcaga aattgtatcc tcgagggttat	1020
aacaggtagg ttgaaggagc atgtaaatgtt cgggactaac ctggaaatgtt ttaggtgttg	1080
gccaatcggtt gatttactca ttacaaatgtt aatcggtga ggaggggttga tgaggattgg	1140
tgaactcttc ctgaatctca gggaaaaatgtt tatttgcaga gtattcaaaa tactgcatt	1200
ttgtggacca atcaaagggg agcttttctt ggatcatggaa gaggtactct tctttggagg	1260
tagcgtgtga aataatgttctt cgcatttttcatcttgcaggatgtt tcttttacct	1320
ctgatcaga ttttccttgcaggatgtt aagggggactt tccttaggaat gaaatgtatccat	1380

cagccagagg ttcccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa	1440
tattttgggt aaaccattt atatcaaaga accttgagtc agatatcctt atcggtttct	1500
ctggctgaag caatgcattt aatgcacaa ttccatctt atgtgcctct cgggcacata	1560
aatatattt gggaatccaa cgaacgacga gctcccat gatctgacag gcgatttcag	1620
gattttctgg acactttgg taggttaga acgtgttagc gtccctgtgt gagaactgac	1680
ggttggatga ggaggaggcc atagccgacg acggaggtt aggctgaggg atggcagact	1740
gggagctcca aactctatag tataccccgtg cgccttcgaa atccgcgcct ccattgtctt	1800
atagtgggtg taaaatgggcc ggaccggggc ggcccagcag gaaaagaagg cgccgactaa	1860
tattaccgcg ccttctttc ctgcgagggc cccgttaggg cccagcgcct tgatttaaag	1920
cctgggtctg ctgttatga ttatctaaa gcagccaaat ctaaagaaaac cggtcccccgg	1980
cactataat tgccctaacaa gtgcgattca ttcatggatc ctttaaactc gagtctagtc	2040
ccgatctagt aacatagatg acaccgcgcg cgataattt taatcgatcc cgccgtatata	2100
tttgtttct atcgcgtatt aatgtataa ttgcggact ctaatcataa aaaccatct	2160
cataaataac gtcatgcatt acatgttaat tattacatgc ttaacgtaat tcaacagaaa	2220
ttatatgata atcatcgaca gaccggcaac aggattcaat ctaaagaaaac tttattgcca	2280
aatgtttgaa cgatcgggga aattcgctcg agttaattaa gcccgcgtt aattaatcg	2340
acgttcttc caaatgaaat gaacttccctt atatagagga agggtttgc gaaggatagt	2400
gggattgtgc gtcatccctt acgtcagttt agatatcaca tcaatccact tgctttgaag	2460
acgtgggtgg aacgttctt tttccacgt agtctctcg tgggtgggggt ccatcttgg	2520
gaccactgtc ggcagaggca tcttgaacga tagcccttcc ttatcgcaat gatggcattt	2580
gtaggtggca ccttctttt ctactgtcct ttgtatgaag tgacagatag ctggcaatg	2640
gaatccgagg aggtttcccg atattaccct ttgttggaaa gtctcaatag ccctttggc	2700
ttctgagact gtatcttga tattcttggaa gtagacgaga gagtgtcggtg ctccaccatg	2760
ttgacgaatt catgggcaga cccgtctgtt cttaagagt gttggcaacc agtaatgaat	2820
aaaaactccc gttttattat atttgatgaa tgctgaaagc ttacattaaat atgtcggtcg	2880
atggcacaa aaaacacacag caaacaatac aggggggttag tcggcgggcg gctaagggtg	2940
gtgctcggcg ggcagaacat cggaaaatca agatctatata gattacact tcctccgtag	3000
gaggaagcac agggggagaa taccacttct ccccccgcga cataatgtaa atgacgcagt	3060
ttgcctcgaa atactccagc tgccctggag tcatttcctt catcaatct tcattccgagt	3120
ttggccgac gatgttaggg ctctgcattt ttgatcttcc atctgtatgac tggatacaga	3180
atccatccat tggagggtcag aaattgcata ctgcagggtt taacaggtag gttgaaggag	3240
catgtaaatc atctacgtat ttgtagatg cgtctcggtt gtatgaagac caatcaacat	3300
tattttggca gtaattatga accccttaggc ttctggccca agtagattt cccgttcttg	3360
ttggccgac gatgttaggg ctctgcattt ttgatcttcc atctgtatgac tggatacaga	3420
atccatccat tggagggtcag aaattgcata ctgcagggtt taacaggtag gttgaaggag	3480
catgtaaatc atctacgtat ttgtagatg cgtctcggtt gtatgaagac caatcaacat	3540
attacaatgtt aaatcagggtt aggggggtgg atgaggattt gtgaaacttctt cctgaatctc	3600
aggaaaaaagc ttatttgccat agtattcaaa atactgcattt ttgttggacc aatcaaagg	3660
gagctcttc tggatcatgg agaggtaatc ttcttggag gtacgtgtt aaataatgtc	3720
tcgcattatt tcatttttag aaggctttt ttcccttacc tctgaatca attttccat	3780
gaagggggac ttcttaggaa tgaaaatgtt cttctcaac acagccagag gttccttgag	3840
aatgtatcc ctcactctgt taactgcattt ggcactctga atatttgggt gaaaccatt	3900
tatataatcc aacccgttgcgat cttccatctt tatgtgcctc tggggcacat agaataatatt tgggaatcca	3960
acgaacgcacg agctcccaga tcatttcgaca ggcgatttca ggattttctg gacactttgg	4020
atagggttagg aacgtgttag cttctgttgc tgagaactga cgggtggatg aggaggaggc	4080
catagccgac gacggaggtt gaggctgagg gatggcagac tggggagctcc aaactctata	4140
gtataccctg ggccttcgaa aatccgcgcg tccattgttct tatgtgggtt gtaaatgggc	4200
cgggccggc cggcccagca gggaaaagaaag ggcgcacta atattaccgc gccttcttt	4260
cctgcggagg cccggggtag ggaccgcgcg ctttgcattt aagcctgggtt ctgtttgtt	4320
tgattttatct aaagcagccc aatctaaaga aaccggccc gggcactata aattgcctaa	4380
caagtgcgtat tcattcatgg atccctttaaa ctgcgttca gagggcccaa ttcgcctat	4440
agtgtgtgtt attacaattc actggccgtc gtttacaac gtgcgtactg gggaaaaccct	4500
ggcgttaccc aacttaatcg cttgcagca catccccctt tgcgcgttgc ggcgtatagc	4560
gaagaggccc gcaccgcgtc cccttcccaa cagttgcgcga gcctataacgt acggcgttt	4620
aaggttaca cctataaaag agagagccgt tttttttttttt tttttttttttt tttttttttttt	4680
attattgaca cgcggggcg acggatgggtt atccccctgg ccagtcacg tctgtgtca	4740
	4800

gataaaagtct cccgtgaact ttacccgggt gtcatatacg gggatgaaag ctggcgcatt	4860
atgaccacccg atatggccag tggccgggtc tccgttatcg gggagaaggat ggctgatctc	4920
agccacccggaaaatgacat caaaaaacgcc attaacctga ttgttctgggg aatataaaatg	4980
tcaggcctga atggcgaatg gacgcgcctt gtagccgcg attaagcgcg cgggtgttgt	5040
ggtaacgcgc acgcgtgaccg ctacacttgc cagcgcctta gcgcgcgc tc ttgcctt	5100
cttcccttcc tttctcgcca ctttcgcggg ctttccccgt caagctctaa atcgggggt	5160
cccttttaggg ttccgattta gagcttaacg gcacctcgac cgcaaaaaac ttgatttggg	5220
tgtatggttca cgtagtgggc catcgccctg atagacgggt ttgcgcctt tgacgttgga	5280
gtccacgttc ttaataatgt gactcttggt ccaaactgga acaacactca accctatcgc	5340
ggctattct tttgatttat aaggatgtt gcccattcg gcctattgggt taaaaaatga	5400
gctgattttaa caaaaaatttt aacaaaatttca agaagaactc gtcaagaagg cgatagaagg	5460
cgatgcgtg cgaatcgggg gcccgcatac cgtaaaagcac gaggaagcgg tcagccatt	5520
cgccgcacag ctcttcagca atatcacggg tagccaaacgc tatgttctga tagcggccg	5580
ccacaccccg ccggccacag tcgatgaatc cagaaaaagcg gcattttcc accatgatat	5640
tcggcaagca ggcacatcgcca tgggtcaca cggatctc gcgcgtcgccg atgctcgct	5700
tgagcctggc gaacagttcg gctggcgcga gcccctgatg ctcttcgtcc agatcatct	5760
gatcgacaaag accgcgttcc atccgactac gtgctcgctc gatgcgtatgt ttgcgttggt	5820
ggtcgaatgg gcaggttagcc ggtcaagcg tatgcagccg cccgatttgc taagccatga	5880
tggatactt ctcgcagga gcaaggtagg atgacaggag atctgcctt ggcacttcgc	5940
ccaatagcag ccagtcctt cccgttccag tgacaacgtc gaggacacgat ggcacaggaa	6000
cgccgcgtg ggccagccac gatagccgcg ctgcctcgatc ttgcgttca ttgcgttgcc	6060
cggaacaggc ggttttgaca aaaagaaccc ggccgcctcg cgctgacacgc cggaacacgg	6120
cggcatcaga gcagccgatt gtctgttgcg cccagtcata gccaatagc ctctccaccc	6180
aaggccgcgg agaacctgcg tgcaatccat ttgttcaat catgcgaaac gatcctcatc	6240
ctgtctttt atcagatctt gatcccctgc gccatcatgat ctttggccgc gagaaagcca	6300
tccagttac tttgcagggc ttcccaacct taccagaggg cggcccaactt ggcaattccg	6360
tttcgttgc tgcataaaa accggccactt ctgcgtatcg ccatgtaaac ccactgcaag	6420
ctacctgtt tctttttgcg cttgcgtttt cccttgcata gatagcccaact tagctgacat	6480
tcatccgggg tcagcaccgt ttctgcggac tggcttcta cgtgaaaagg atcttaggtga	6540
agatccttt tgataatctc atgaccaaaa tcccttaacg tgagtttgc ttccactgag	6600
cgtcagaccc cgtagaaaaag atcaaaggat ttcttgaga tccctttttt ctgcgtgtaa	6660
tctgtgttt gcaaacaaaa aaaccaacccg taccagccgt ggtttgtttt ccggatcaag	6720
agctaccaac tctttttccg aaggttaactg gcttcagcag agcgcagata ccaaataactg	6780
tccttctagt gttagccgt tagtgcacc acttcagaa ctctgttagca ccgcctacat	6840
acctcgctc gctaattctg ttaccagttt ctgcgtcccg tggcgataag tctgttctta	6900
ccgggttggc ctcaagacga tagttaccgg ataaggcgca ggggtcgccg tgaacgggg	6960
tttcgtgcac acagccacgc ttggagcgaa cgacccatcac cgaactgaga tacctacagc	7020
gtgagctatg agaaagcgcc acgcttcccg aaggagaaaa ggcggacagg tatccggtaa	7080
gcggcagggt cggaacaggaa gagcgcacga gggagcttcc agggggaaac gcctggatc	7140
tttatagtcc tgcgtgggtt cggccacccctt gacttgagcg tgcattttt tgatgtctgt	7200
caggggggcg gagctatgg aaaaaacgcga gcaacgcgcgc ctttttacgg ttccctgggt	7260
tttgcgtgcct ttttgcac atgttcttgc ctgcgttatac ccctgattct gtggataacc	7320
gtattacccgc ctggagttga gctgataccg ctgcgcgcgag cggaaacgcacc gagcgcagcg	7380
agtcagttagg cgaggaagcg gaag	7404